PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6: C12N 15/12, C07K 14/475, A61K 38/18, C12Q 1/68, C07K 16/22 // C12N 15/62, 9/00, 15/11

A1

(11) International Publication Number: WO 98/41634

(43) International Publication Date: 24 September 1998 (24.09.98)

(21) International Application Number: PCT/US98/05700

(22) International Filing Date: 20 March 1998 (20.03.98)

(30) Priority Data:
08/822,259 20 March 1997 (20.03.97) US

(63) Related by Continuation (CON) or Continuation-in-Part (CIP) to Earlier Application
US 08/822,259 (CIP)
Filed on 20 March 1997 (20.03.97)

(71) Applicant *(for all designated States except US)*: INCYTE PHARMACEUTICALS, INC. [US/US]; 3174 Porter Drive, Palo Alto, CA 94304 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: WILDE, Craig, G. [US/US]; 1239 Mandarin Drive, Sunnyvale, CA 94087 (US). GOLI, Surya, K. [IN/US]; 3311 Cantamar Court, San Jose, CA 9513 (US).

(74) Agent: BILLINGS, Lucy, J.; Incyte Pharmaceuticals, Inc., 3174 Porter Drive, Palo Alto, CA 94304 (US).

(81) Designated States: AT, AU, BR, CA, CH, CN, DE, DK, ES, FI, GB, IL, JP, KR, MX, NO, NZ, RU, SE, SG, US, ARIPO patent (GH, GM, KE, LS, MW, SD, SZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, ML, MR, NE, SN, TD, TG).

Published
With international search report.

(54) Title: HUMAN GROWTH AND DIFFERENTIATION PROTEIN

(57) Abstract

The present invention provides a human growth and differentiation protein (GRODIF) and polynucleotides which identify and encode GRODIF. The invention also provides expression vectors, host cells, agonists, antagonists and antibodies. The invention also provides methods to treat disorders associated with cell proliferation or expression of GRODIF.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AL	Albania	**ES**	Spain	**LS**	Lesotho	**SI**	Slovenia
AM	Armenia	**FI**	Finland	**LT**	Lithuania	**SK**	Slovakia
AT	Austria	**FR**	France	**LU**	Luxembourg	**SN**	Senegal
AU	Australia	**GA**	Gabon	**LV**	Latvia	**SZ**	Swaziland
AZ	Azerbaijan	**GB**	United Kingdom	**MC**	Monaco	**TD**	Chad
BA	Bosnia and Herzegovina	**GE**	Georgia	**MD**	Republic of Moldova	**TG**	Togo
BB	Barbados	**GH**	Ghana	**MG**	Madagascar	**TJ**	Tajikistan
BE	Belgium	**GN**	Guinea	**MK**	The former Yugoslav Republic of Macedonia	**TM**	Turkmenistan
BF	Burkina Faso	**GR**	Greece			**TR**	Turkey
BG	Bulgaria	**HU**	Hungary	**ML**	Mali	**TT**	Trinidad and Tobago
BJ	Benin	**IE**	Ireland	**MN**	Mongolia	**UA**	Ukraine
BR	Brazil	**IL**	Israel	**MR**	Mauritania	**UG**	Uganda
BY	Belarus	**IS**	Iceland	**MW**	Malawi	**US**	United States of America
CA	Canada	**IT**	Italy	**MX**	Mexico	**UZ**	Uzbekistan
CF	Central African Republic	**JP**	Japan	**NE**	Niger	**VN**	Viet Nam
CG	Congo	**KE**	Kenya	**NL**	Netherlands	**YU**	Yugoslavia
CH	Switzerland	**KG**	Kyrgyzstan	**NO**	Norway	**ZW**	Zimbabwe
CI	Côte d'Ivoire	**KP**	Democratic People's Republic of Korea	**NZ**	New Zealand		
CM	Cameroon			**PL**	Poland		
CN	China	**KR**	Republic of Korea	**PT**	Portugal		
CU	Cuba	**KZ**	Kazakstan	**RO**	Romania		
CZ	Czech Republic	**LC**	Saint Lucia	**RU**	Russian Federation		
DE	Germany	**LI**	Liechtenstein	**SD**	Sudan		
DK	Denmark	**LK**	Sri Lanka	**SE**	Sweden		
EE	Estonia	**LR**	Liberia	**SG**	Singapore		

HUMAN GROWTH AND DIFFERENTIATION PROTEIN

TECHNICAL FIELD

This invention relates to nucleic acid and amino acid sequences of a novel human growth and differentiation protein and to the use of these sequences in the diagnosis, prevention, and treatment of conditions and disorders involving cell proliferation.

BACKGROUND ART

The transforming growth factor-beta (TGF-B) superfamily encompasses a group of structurally related proteins that control cell proliferation, differentiation, and other functions in many cell types (Roberts, A. and Sporn, M. (1990) In: Peptide Growth Factors and Their Receptors. Handbook of Experimental Pharmacology, 95:419-475; Springer-Verlag, Berlin). The prototypical molecules, TGF-B1, 2 and 3, can influence a wide variety of proliferation and differentiation processes including adipogenesis, embryogenesis, chondrogenesis, hematopoiesis, and epithelial differentiation (Massague, J. (1987) Cell 49:437-438).

Depending on the cell lineage, TGF-B can have a stimulatory or inhibitory effect on cell proliferation or differentiation. For instance, TGF-B has been shown to have a wide range of immunomodulatory activities including potent immunosuppressive effects on B and T-cell proliferation and inhibitory effects on the growth of early hematopoietic progenitor cells (Palladino, M. et al. (1990) Ann. N.Y. Acad. Sci. 593:181-187; Moore, M. (1991) Blood 78:1-19). Transgenic mice expressing a TGF-B1 null mutation exhibit a phenotype with excessive inflammatory response and an early death (Kulkarni, A. (1993) Proc. Natl. Acad. Sci. 90:770-774). Neutralizing antibodies to TGF-B have been shown to release hematopoietic stem cells from the inhibitory effect of endogenous TGF-B and permit enhanced proliferation of stem cells in long term bone marrow cultures (Waegell, W. et al. (1994) Exp. Hematol. 22:1051-1057). In contrast, exogenous TGF-B has been shown to enhance the differentiation of mesenchymal cells such as fibroblasts in culture and increase their synthesis of matrix molecules (Bassols, A. and Massague, J. (1988) J. Biol. Chem. 263:3039-3045). Similarly, treatment with TGF-B has proven of value in enhancing wound strength and healing in several models of tissue repair (Mustoe, T. et al. (1987) Science 237:1333-1336).

Other molecules in the TGF-B superfamily include: Mullerian inhibiting substance, produced by the testis and required for normal male sexual development (Behringer, R. et al. (1990) Nature 345:167-170); inhibins and activins, which inhibit or stimulate the secretion of follitropin by the pituitary or induce erythroid differentiation in the case of activin (Shiozaki, M. et al. (1992) Proc. Natl. Acad. Sci. U.S.A. 89:1553-1556); the bone morphogenetic proteins

(BMPs, VGRs, and Ops), which induce cartilage and bone formation and are involved in skeletal development (Sampath, T. et al. (1990) J. Biol. Chem. 265:1398-13205); and the Xenopus Vg-1 egg gene product which can induce the formation of mesoderm and anterior structures in the amphibian embryo (Mayo, K. et al. (1986) Proc. Natl. Acad. Sci. U.S.A. 83:5849-5853). A more recently identified novel family member, growth and differentiation factor-1 (GDF-1), has been shown to be expressed almost exclusively in the nervous system (Lee, S. (1990) Mol. Endocrinol. 4:1034-1040).

Each member of the TGF-B superfamily is thought to be synthesized as part of a larger precursor protein that undergoes proteolytic cleavage to produce an active C-terminus fragment. The C-terminus amino acid homologies range from 70-90% within the different subgroups to 20-50% between more distantly related family members. The functional protein in each instance appears to be a disulfide-linked dimer. Homodimeric complexes appear to be most common, however heterodimers between different proteins which confer distinct biological properties have been described (Cheifetz, S. et al. (1987) Cell 48:409-415).

Techniques using degenerate oligonucleotides from the highly conserved regions among known TGF-B subgroup members as hybridization probes or PCR primers have recently identified new mammalian members of the superfamily including GDF-3 and GDF-9 (McPherron, A. and Lee, S. (1993) J. Biol. Chem. 268(5):3444-3449). The COOH-terminal portion of the GDF-3 molecule showed significant homology (53-57%) to known members of the TGF-B family including Vgr-1 and BMP-2. The GDF-9 sequence diverged significantly from all other members. Unlike all previously described members of the TGF-B family, both GDF-3 and GDF-9 lack the single conserved cysteine residue believed to form the disulfide linkage between dimeric subunits. The expression pattern of murine GDF-3, by Northern analysis of adult mouse tissues showed the highest levels of mRNA in thymus, spleen, bone marrow, and adipose tissue. No transcripts were detected in brain, liver, kidney, muscle, and testis. Hybridization with a GDF-9 probe against a similar panel of tissue specific RNAs detected expression exclusively in the mouse ovary. Recently a human GDF-3-like protein has been described with significant homologies to the murine sequence (Lee, McPherron, WO94/15965). This pattern of expression suggests that GDF-3, GDF-9 and GDF-like molecules may play a critical role in the regulation and differentiation of both the immuno-hematopoietic, connective tissue, and reproductive systems.

Molecules related to human and mouse GDF proteins satisfy a need in the art by providing new compositions useful in diagnosing and treating conditions and disorders involving

cell proliferation.

DISCLOSURE OF THE INVENTION

The present invention features a novel TGF-B superfamily protein hereinafter designated GRODIF and characterized as having similarity to the murine GDF-3 protein.

Accordingly, the invention features a substantially purified GRODIF having the amino acid sequence shown in SEQ ID NO:1.

One aspect of the invention features isolated and substantially purified polynucleotides that encode GRODIF. In a particular aspect, the polynucleotide is the nucleotide sequence of SEQ ID NO:2.

The invention also relates to a polynucleotide sequence comprising the complement of SEQ ID NO:2 or variants thereof. In addition, the invention features polynucleotide sequences which hybridize under stringent conditions to SEQ ID NO:2.

The invention additionally features nucleic acid sequences encoding polypeptides, oligonucleotides, peptide nucleic acids (PNA), fragments, portions or antisense molecules thereof, and expression vectors and host cells comprising polynucleotides that encode GRODIF. The present invention also features antibodies which bind specifically to GRODIF, and pharmaceutical compositions comprising substantially purified GRODIF. The invention also features agonists and antagonists of GRODIF. The invention also features methods for using the protein, agonists and antagonists to treat conditions involving cell proliferation.

BRIEF DESCRIPTION OF DRAWINGS

Figures 1A, 1B, 1C, and 1D show the amino acid sequence (SEQ ID NO:1) and nucleic acid sequence (SEQ ID NO:2) of GRODIF. The alignment was produced using MacDNASIS PRO™ software (Hitachi Software Engineering Co. Ltd. San Bruno, CA).

Figures 2A, 2B and 2C show the amino acid sequence alignments among GRODIF (SEQ ID NO:1), a human GDF protein (WO94/15965; SEQ ID NO:3), and murine GDF-3 (GI 567205; SEQ ID NO:4). The alignment was produced using the multisequence alignment program of DNASTAR™ software (DNASTAR Inc, Madison WI).

Figures 3A and 3B show the hydrophobicity plots (MacDNASIS PRO software) for GRODIF, SEQ ID NO:1 and for murine GDF-3 protein, SEQ ID NO:4, respectively. The positive X axis reflects amino acid position, and the negative Y axis, hydrophobicity.

MODES FOR CARRYING OUT THE INVENTION

Before the present proteins, nucleotide sequences, and methods are described, it is understood that this invention is not limited to the particular methodology, protocols, cell lines,

vectors, and reagents described as these may vary. It is also to be understood that the terminology used herein is for the purpose of describing particular embodiments only, and is not intended to limit the scope of the present invention which will be limited only by the appended claims.

It must be noted that as used herein and in the appended claims, the singular forms "a", "an", and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, reference to "a host cell" includes a plurality of such host cells, reference to the "antibody" is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any methods and materials similar or equivalent to those described herein can be used in the practice or testing of the present invention, the preferred methods, devices, and materials are now described. All publications mentioned herein are incorporated herein by reference for the purpose of describing and disclosing the cell lines, vectors, and methodologies which are reported in the publications which might be used in connection with the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

DEFINITIONS

"Nucleic acid sequence" as used herein refers to an oligonucleotide, nucleotide, or polynucleotide, and fragments or portions thereof, and to DNA or RNA of genomic or synthetic origin which may be single- or double-stranded, and represent the sense or antisense strand. Similarly, "amino acid sequence" as used herein refers to an oligopeptide, peptide, polypeptide, or protein sequence, and fragments or portions thereof, and to naturally occurring or synthetic molecules.

Where "amino acid sequence" is recited herein to refer to an amino acid sequence of a naturally occurring protein molecule, "amino acid sequence" and like terms, such as "polypeptide" or "protein" are not meant to limit the amino acid sequence to the complete, native amino acid sequence associated with the recited protein molecule.

"Peptide nucleic acid", as used herein, refers to a molecule which comprises an oligomer to which an amino acid residue, such as lysine, and an amino group have been added. These small molecules, also designated anti-gene agents, stop transcript elongation by binding to their complementary strand of nucleic acid (Nielsen, P.E. et al. (1993) Anticancer Drug Des. 8:53-63).

GRODIF, as used herein, refers to the amino acid sequences of substantially purified

GRODIF obtained from any species, particularly mammalian, including bovine, ovine, porcine, murine, equine, and preferably human, from any source whether natural, synthetic, semi-synthetic, or recombinant.

"Consensus", as used herein, refers to a nucleic acid sequence which has been resequenced to resolve uncalled bases, or which has been extended using XL-PCR™ (Perkin Elmer, Norwalk, CT) in the 5' and/or the 3' direction and resequenced, or which has been assembled from the overlapping sequences of more than one Incyte clone using the GELVIEW™ Fragment Assembly system (GCG, Madison, WI), or which has been both extended and assembled.

A "variant" of GRODIF, as used herein, refers to an amino acid sequence that is altered by one or more amino acids. The variant may have "conservative" changes, wherein a substituted amino acid has similar structural or chemical properties, e.g., replacement of leucine with isoleucine. More rarely, a variant may have "nonconservative" changes, e.g., replacement of a glycine with a tryptophan. Similar minor variations may also include amino acid deletions or insertions, or both. Guidance in determining which amino acid residues may be substituted, inserted, or deleted without abolishing biological or immunological activity may be found using computer programs well known in the art, for example, DNASTAR software.

A "deletion", as used herein, refers to a change in either amino acid or nucleotide sequence in which one or more amino acid or nucleotide residues, respectively, are absent.

An "insertion" or "addition", as used herein, refers to a change in an amino acid or nucleotide sequence resulting in the addition of one or more amino acid or nucleotide residues, respectively, as compared to the naturally occurring molecule.

A "substitution", as used herein, refers to the replacement of one or more amino acids or nucleotides by different amino acids or nucleotides, respectively.

The term "biologically active", as used herein, refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" refers to the capability of the natural, recombinant, or synthetic GRODIF, or any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

The term "agonist", as used herein, refers to a molecule which, when bound to GRODIF, causes a change in GRODIF which modulates the activity of GRODIF. Agonists may include proteins, nucleic acids, carbohydrates, or any other molecules which bind to GRODIF.

The terms "antagonist" or "inhibitor", as used herein, refer to a molecule which, when

bound to GRODIF, blocks or modulates the biological or immunological activity of GRODIF. Antagonists and inhibitors may include proteins, nucleic acids, carbohydrates, or any other molecules which bind to GRODIF.

The term "modulate", as used herein, refers to a change or an alteration in the biological activity of GRODIF. Modulation may be an increase or a decrease in protein activity, a change in binding characteristics, or any other change in the biological, functional or immunological properties of GRODIF.

The term "mimetic", as used herein, refers to a molecule, the structure of which is developed from knowledge of the structure of GRODIF or portions thereof and, as such, is able to effect some or all of the actions of [Human GDF-3-like protein]-like molecules.

The term "derivative", as used herein, refers to the chemical modification of a nucleic acid encoding GRODIF or the encoded GRODIF. Illustrative of such modifications would be replacement of hydrogen by an alkyl, acyl, or amino group. A nucleic acid derivative would encode a polypeptide which retains essential biological characteristics of the natural molecule.

The term "substantially purified", as used herein, refers to nucleic or amino acid sequences that are removed from their natural environment, isolated or separated, and are at least 60% free, preferably 75% free, and most preferably 90% free from other components with which they are naturally associated.

"Amplification" as used herein refers to the production of additional copies of a nucleic acid sequence and is generally carried out using polymerase chain reaction (PCR) technologies well known in the art (Dieffenbach, C.W. and G.S. Dveksler (1995) PCR Primer, a Laboratory Manual, Cold Spring Harbor Press, Plainview, NY).

The term "hybridization", as used herein, refers to any process by which a strand of nucleic acid binds with a complementary strand through base pairing.

The term "hybridization complex", as used herein, refers to a complex formed between two nucleic acid sequences by virtue of the formation of hydrogen binds between complementary G and C bases and between complementary A and T bases; these hydrogen bonds may be further stabilized by base stacking interactions. The two complementary nucleic acid sequences hydrogen bond in an antiparallel configuration. A hybridization complex may be formed in solution (e.g., C_0t or R_0t analysis) or between one nucleic acid sequence present in solution and another nucleic acid sequence immobilized on a solid support (e.g., membranes, filters, chips, pins or glass slides to which cells have been fixed for in situ hybridization).

The terms "complementary" or "complementarity", as used herein, refer to the natural

binding of polynucleotides under permissive salt and temperature conditions by base-pairing. For example, for the sequence "A-G-T" binds to the complementary sequence "T-C-A". Complementarity between two single-stranded molecules may be "partial", in which only some of the nucleic acids bind, or it may be complete when total complementarity exists between the single stranded molecules. The degree of complementarity between nucleic acid strands has significant effects on the efficiency and strength of hybridization between nucleic acid strands. This is of particular importance in amplification reactions, which depend upon binding between nucleic acids strands.

The term "homology", as used herein, refers to a degree of complementarity. There may be partial homology or complete homology (i.e., identity). A partially complementary sequence is one that at least partially inhibits an identical sequence from hybridizing to a target nucleic acid; it is referred to using the functional term "substantially homologous." The inhibition of hybridization of the completely complementary sequence to the target sequence may be examined using a hybridization assay (Southern or northern blot, solution hybridization and the like) under conditions of low stringency. A substantially homologous sequence or probe will compete for and inhibit the binding (i.e., the hybridization) of a completely homologous sequence or probe to the target sequence under conditions of low stringency. This is not to say that conditions of low stringency are such that non-specific binding is permitted; low stringency conditions require that the binding of two sequences to one another be a specific (i.e., selective) interaction. The absence of non-specific binding may be tested by the use of a second target sequence which lacks even a partial degree of complementarity (e.g., less than about 30% identity); in the absence of non-specific binding, the probe will not hybridize to the second non-complementary target sequence.

As known in the art, numerous equivalent conditions may be employed to comprise either low or high stringency conditions. Factors such as the length and nature (DNA, RNA, base composition) of the sequence, nature of the target (DNA, RNA, base composition, presence in solution or immobilization, etc.), and the concentration of the salts and other components (e.g., the presence or absence of formamide, dextran sulfate and/or polyethylene glycol) are considered and the hybridization solution may be varied to generate conditions of either low or high stringency different from, but equivalent to, the above listed conditions.

The term "stringent conditions", as used herein, is the "stringency" which occurs within a range from about Tm-5°C (5°C below the melting temperature (Tm) of the probe) to about 20°C to 25°C below Tm. As will be understood by those of skill in the art, the stringency of

hybridization may be altered in order to identify or detect identical or related polynucleotide sequences.

The term "antisense", as used herein, refers to nucleotide sequences which are complementary to a specific DNA or RNA sequence. The term "antisense strand" is used in reference to a nucleic acid strand that is complementary to the "sense" strand. Antisense molecules may be produced by any method, including synthesis by ligating the gene(s) of interest in a reverse orientation to a viral promoter which permits the synthesis of a complementary strand. Once introduced into a cell, this transcribed strand combines with natural sequences produced by the cell to form duplexes. These duplexes then block either the further transcription or translation. In this manner, mutant phenotypes may be generated. The designation "negative" is sometimes used in reference to the antisense strand, and "positive" is sometimes used in reference to the sense strand.

The term "portion", as used herein, with regard to a protein (as in "a portion of a given protein") refers to fragments of that protein. The fragments may range in size from four amino acid residues to the entire amino acid sequence minus one amino acid. Thus, a protein "comprising at least a portion of the amino acid sequence of SEQ ID NO:1" encompasses the full-length human GRODIF and fragments thereof.

"Transformation", as defined herein, describes a process by which exogenous DNA enters and changes a recipient cell. It may occur under natural or artificial conditions using various methods well known in the art. Transformation may rely on any known method for the insertion of foreign nucleic acid sequences into a prokaryotic or eukaryotic host cell. The method is selected based on the host cell being transformed and may include, but is not limited to, viral infection, electroporation, lipofection, and particle bombardment. Such "transformed" cells include stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome. They also include cells which transiently express the inserted DNA or RNA for limited periods of time.

The term "antigenic determinant", as used herein, refers to that portion of a molecule that makes contact with a particular antibody (i.e., an epitope). When a protein or fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to a given region or three-dimensional structure on the protein; these regions or structures are referred to as antigenic determinants. An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

The terms "specific binding" or "specifically binding", as used herein, in reference to the interaction of an antibody and a protein or peptide, mean that the interaction is dependent upon the presence of a particular structure (i.e., the antigenic determinant or epitope) on the protein; in other words, the antibody is recognizing and binding to a specific protein structure rather than to proteins in general. For example, if an antibody is specific for epitope "A", the presence of a protein containing epitope A (or free, unlabeled A) in a reaction containing labeled "A" and the antibody will reduce the amount of labeled A bound to the antibody.

The term "sample", as used herein, is used in its broadest sense. A biological sample suspected of containing nucleic acid encoding GRODIF or fragments thereof may comprise a cell, chromosomes isolated from a cell (e.g., a spread of metaphase chromosomes), genomic DNA (in solution or bound to a solid support such as for Southern analysis), RNA (in solution or bound to a solid support such as for northern analysis), cDNA (in solution or bound to a solid support), an extract from cells or a tissue, and the like.

The term "correlates with expression of a polynucleotide", as used herein, indicates that the detection of the presence of ribonucleic acid that is similar to SEQ ID NO:2 by northern analysis is indicative of the presence of mRNA encoding GRODIF in a sample and thereby correlates with expression of the transcript from the polynucleotide encoding the protein.

"Alterations" in the polynucleotide of SEQ ID NO: 2, as used herein, comprise any alteration in the sequence of polynucleotides encoding GRODIF including deletions, insertions, and point mutations that may be detected using hybridization assays. Included within this definition is the detection of alterations to the genomic DNA sequence which encodes GRODIF (e.g., by alterations in the pattern of restriction fragment length polymorphisms capable of hybridizing to SEQ ID NO:2), the inability of a selected fragment of SEQ ID NO: 2 to hybridize to a sample of genomic DNA (e.g., using allele-specific oligonucleotide probes), and improper or unexpected hybridization, such as hybridization to a locus other than the normal chromosomal locus for the polynucleotide sequence encoding GRODIF (e.g., using fluorescent in situ hybridization [FISH] to metaphase chromosomes spreads).

As used herein, the term "antibody" refers to intact molecules as well as fragments thereof, such as Fa, $F(ab')_2$, and Fv, which are capable of binding the epitopic determinant. Antibodies that bind GRODIF polypeptides can be prepared using intact polypeptides or fragments containing small peptides of interest as the immunizing antigen. The polypeptide or peptide used to immunize an animal can be derived from the transition of RNA or synthesized chemically, and can be conjugated to a carrier protein, if desired. Commonly used carriers that

are chemically coupled to peptides include bovine serum albumin and thyroglobulin. The coupled peptide is then used to immunize the animal (e.g., a mouse, a rat, or a rabbit).

The term "humanized antibody", as used herein, refers to antibody molecules in which amino acids have been replaced in the non-antigen binding regions in order to more closely resemble a human antibody, while still retaining the original binding ability.

THE INVENTION

The invention is based on the discovery of a novel human growth and differentiation protein, (GRODIF), the polynucleotides encoding GRODIF, and the use of these compositions for the diagnosis, prevention, or treatment of disorders associated with cell proliferation.

Nucleic acids encoding the human GRODIF of the present invention were first identified in Incyte Clone 266285 from the hNT-2 cDNA library (HNT2NOT01) through a computer-generated search for amino acid sequence alignments. A consensus sequence, SEQ ID NO:2, was derived from the following overlapping and/or extended nucleic acid sequences: Incyte Clones 2342537 and 2346385 (TESTTUT02), and 266285 (HNT2NOT01).

In one embodiment, the invention encompasses a polypeptide comprising the amino acid sequence of SEQ ID NO:1, as shown in Figures 1A, 1B, 1C and 1D. GRODIF is 364 amino acids in length, has two potential N-glycosylation sites, N_{112} and N_{306}, and one potential myristylation site, G_{77}. As shown in Figures 2A, 2B and 2C, GRODIF displays the TGF-beta family signature from I_{282}-C_{297}; and except for the absence of a cysteine at position 328, has a conserved cysteine pattern consistent with assignment to TGF-beta family superfamily. GRODIF has chemical and structural homology with murine GDF-3 protein (GI 567205; SEQ ID NO:4) and an identical human GDF protein (WO94/15965; SEQ ID NO:3). In particular, GRODIF shares 69% identity with the murine GDF-3 protein and the human GDF protein. As illustrated by Figures 3A and 3B, although GRODIF and murine GDF-3 protein have rather similar hydrophobicity plots. The calculated isoelectric points are 7.83 for GRODIF and 9.49 for the murine GDF-3 protein.

The invention also encompasses the formation of heterodimeric polypeptides between subunits of GRODIF and of other members of the TGF-B superfamily including, but not limited to, TGF isoforms; GDF isoforms, Mullerian inhibiting substance, inhibin, activin, BMPs, and Vg-1.

The invention also encompasses GRODIF variants. A preferred GRODIF variant is one having at least 80%, and more preferably 90%, amino acid sequence identity to the GRODIF amino acid sequence (SEQ ID NO:1). A most preferred GRODIF variant is one having at least

95% amino acid sequence identity to SEQ ID NO:1.

The invention also encompasses polynucleotides which encode GRODIF. Accordingly, any nucleic acid sequence which encodes the amino acid sequence of GRODIF can be used to generate recombinant molecules which express GRODIF. In a particular embodiment, the invention encompasses the polynucleotide comprising the nucleic acid sequence of SEQ ID NO:2 as shown in Figures 1A, 1B, 1C and 1D.

It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of nucleotide sequences encoding GRODIF, some bearing minimal homology to the nucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of nucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the nucleotide sequence of naturally occurring GRODIF, and all such variations are to be considered as being specifically disclosed.

Although nucleotide sequences which encode GRODIF and its variants are preferably capable of hybridizing to the nucleotide sequence of the naturally occurring GRODIF under appropriately selected conditions of stringency, it may be advantageous to produce nucleotide sequences encoding GRODIF or its derivatives possessing a substantially different codon usage. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding GRODIF and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

The invention also encompasses production of DNA sequences, or portions thereof, which encode GRODIF and its derivatives, entirely by synthetic chemistry. After production, the synthetic sequence may be inserted into any of the many available expression vectors and cell systems using reagents that are well known in the art at the time of the filing of this application. Moreover, synthetic chemistry may be used to introduce mutations into a sequence encoding GRODIF or any portion thereof.

Also encompassed by the invention are polynucleotide sequences that are capable of hybridizing to the claimed nucleotide sequences, and in particular, those shown in SEQ ID NO:2, under various conditions of stringency. Hybridization conditions are based on the melting

temperature (Tm) of the nucleic acid binding complex or probe, as taught in Wahl, G.M. and S.L. Berger (1987; Methods Enzymol. 152:399-407) and Kimmel, A.R. (1987; Methods Enzymol. 152:507-511), and may be used at a defined stringency.

Altered nucleic acid sequences encoding GRODIF which are encompassed by the invention include deletions, insertions, or substitutions of different nucleotides resulting in a polynucleotide that encodes the same or a functionally equivalent GRODIF. The encoded protein may also contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent GRODIF. Deliberate amino acid substitutions may be made on the basis of similarity in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues as long as the biological activity of GRODIF is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid; positively charged amino acids may include lysine and arginine; and amino acids with uncharged polar head groups having similar hydrophilicity values may include leucine, isoleucine, and valine; glycine and alanine; asparagine and glutamine; serine and threonine; phenylalanine and tyrosine.

Also included within the scope of the present invention are alleles of the genes encoding GRODIF. As used herein, an "allele" or "allelic sequence" is an alternative form of the gene which may result from at least one mutation in the nucleic acid sequence. Alleles may result in altered mRNAs or polypeptides whose structure or function may or may not be altered. Any given gene may have none, one, or many allelic forms. Common mutational changes which give rise to alleles are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

Methods for DNA sequencing which are well known and generally available in the art may be used to practice any embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of DNA polymerase I, Sequenase® (US Biochemical Corp, Cleveland, OH), Taq polymerase (Perkin Elmer), thermostable T7 polymerase (Amersham, Chicago, IL), or combinations of recombinant polymerases and proofreading exonucleases such as the ELONGASE Amplification System marketed by Gibco BRL (Gaithersburg, MD). Preferably, the process is automated with machines such as the Hamilton Micro Lab 2200 (Hamilton, Reno, NV), Peltier Thermal Cycler (PTC200; MJ Research, Watertown, MA) and the ABI 377 DNA sequencers (Perkin Elmer).

The nucleic acid sequences encoding GRODIF may be extended utilizing a partial

nucleotide sequence and employing various methods known in the art to detect upstream sequences such as promoters and regulatory elements. For example, one method which may be employed, "restriction-site" PCR, uses universal primers to retrieve unknown sequence adjacent to a known locus (Sarkar, G. (1993) PCR Methods Applic. 2:318-322). In particular, genomic DNA is first amplified in the presence of primer to linker sequence and a primer specific to the known region. The amplified sequences are then subjected to a second round of PCR with the same linker primer and another specific primer internal to the first one. Products of each round of PCR are transcribed with an appropriate RNA polymerase and sequenced using reverse transcriptase.

Inverse PCR may also be used to amplify or extend sequences using divergent primers based on a known region (Triglia, T. et al. (1988) Nucleic Acids Res. 16:8186). The primers may be designed using OLIGO 4.06 Primer Analysis software (National Biosciences Inc., Plymouth, MN), or another appropriate program, to be 22-30 nucleotides in length, to have a GC content of 50% or more, and to anneal to the target sequence at temperatures about 68°-72° C. The method uses several restriction enzymes to generate a suitable fragment in the known region of a gene. The fragment is then circularized by intramolecular ligation and used as a PCR template.

Another method which may be used is capture PCR which involves PCR amplification of DNA fragments adjacent to a known sequence in human and yeast artificial chromosome DNA (Lagerstrom, M. et al. (1991) PCR Methods Applic. 1:111-119). In this method, multiple restriction enzyme digestions and ligations may also be used to place an engineered double-stranded sequence into an unknown portion of the DNA molecule before performing PCR.

Another method which may be used to retrieve unknown sequences is that of Parker, J.D. et al. (1991; Nucleic Acids Res. 19:3055-3060). Additionally, one may use PCR, nested primers, and PromoterFinder™ libraries to walk in genomic DNA (Clontech, Palo Alto, CA). This process avoids the need to screen libraries and is useful in finding intron/exon junctions.

When screening for full-length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. Also, random-primed libraries are preferable, in that they will contain more sequences which contain the 5' regions of genes. Use of a randomly primed library may be especially preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into the 5' and 3' non-transcribed regulatory regions.

Capillary electrophoresis systems which are commercially available may be used to

analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different fluorescent dyes (one for each nucleotide) which are laser activated, and detection of the emitted wavelengths by a charge coupled devise camera. Output/light intensity may be converted to electrical signal using appropriate software (e.g. Genotyper™ and Sequence Navigator™, Perkin Elmer) and the entire process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for the sequencing of small pieces of DNA which might be present in limited amounts in a particular sample.

In another embodiment of the invention, polynucleotide sequences or fragments thereof which encode GRODIF, or fusion proteins or functional equivalents thereof, may be used in recombinant DNA molecules to direct expression of GRODIF in appropriate host cells. Due to the inherent degeneracy of the genetic code, other DNA sequences which encode substantially the same or a functionally equivalent amino acid sequence may be produced and these sequences may be used to clone and express GRODIF.

As will be understood by those of skill in the art, it may be advantageous to produce GRODIF-encoding nucleotide sequences possessing non-naturally occurring codons. For example, codons preferred by a particular prokaryotic or eukaryotic host can be selected to increase the rate of protein expression or to produce a recombinant RNA transcript having desirable properties, such as a half-life which is longer than that of a transcript generated from the naturally occurring sequence.

The nucleotide sequences of the present invention can be engineered using methods generally known in the art in order to alter GRODIF encoding sequences for a variety of reasons, including but not limited to, alterations which modify the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, site-directed mutagenesis may be used to insert new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, or introduce mutations, and so forth.

In another embodiment of the invention, natural, modified, or recombinant nucleic acid sequences encoding GRODIF may be ligated to a heterologous sequence to encode a fusion protein. For example, to screen peptide libraries for inhibitors of GRODIF activity, it may be useful to encode a chimeric GRODIF protein that can be recognized by a commercially available antibody. A fusion protein may also be engineered to contain a cleavage site located between the

GRODIF encoding sequence and the heterologous protein sequence, so that GRODIF may be cleaved and purified away from the heterologous moiety.

In another embodiment, sequences encoding GRODIF may be synthesized, in whole or in part, using chemical methods well known in the art (see Caruthers, M.H. et al. (1980) Nucl. Acids Res. Symp. Ser. 215-223, Horn, T. et al. (1980) Nucl. Acids Res. Symp. Ser. 225-232). Alternatively, the protein itself may be produced using chemical methods to synthesize the amino acid sequence of GRODIF, or a portion thereof. For example, peptide synthesis can be performed using various solid-phase techniques (Roberge, J.Y. et al. (1995) Science 269:202-204) and automated synthesis may be achieved, for example, using the ABI 431A Peptide Synthesizer (Perkin Elmer).

The newly synthesized peptide may be substantially purified by preparative high performance liquid chromatography (e.g., Creighton, T. (1983) Proteins, Structures and Molecular Principles, WH Freeman and Co., New York, NY). The composition of the synthetic peptides may be confirmed by amino acid analysis or sequencing (e.g., the Edman degradation procedure; Creighton, supra). Additionally, the amino acid sequence of GRODIF, or any part thereof, may be altered during direct synthesis and/or combined using chemical methods with sequences from other proteins, or any part thereof, to produce a variant polypeptide.

In order to express a biologically active GRODIF, the nucleotide sequences encoding GRODIF or functional equivalents, may be inserted into appropriate expression vector, i.e., a vector which contains the necessary elements for the transcription and translation of the inserted coding sequence.

Methods which are well known to those skilled in the art may be used to construct expression vectors containing sequences encoding GRODIF and appropriate transcriptional and translational control elements. These methods include in vitro recombinant DNA techniques, synthetic techniques, and in vivo genetic recombination. Such techniques are described in Sambrook, J. et al. (1989) Molecular Cloning, A Laboratory Manual, Cold Spring Harbor Press, Plainview, NY, and Ausubel, F.M. et al. (1989) Current Protocols in Molecular Biology, John Wiley & Sons, New York, NY.

A variety of expression vector/host systems may be utilized to contain and express sequences encoding GRODIF. These include, but are not limited to, microorganisms such as bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with virus expression vectors (e.g., baculovirus); plant cell systems transformed with virus expression

vectors (e.g., cauliflower mosaic virus, CaMV; tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems.

The "control elements" or "regulatory sequences" are those non-translated regions of the vector--enhancers, promoters, 5' and 3' untranslated regions--which interact with host cellular proteins to carry out transcription and translation. Such elements may vary in their strength and specificity. Depending on the vector system and host utilized, any number of suitable transcription and translation elements, including constitutive and inducible promoters, may be used. For example, when cloning in bacterial systems, inducible promoters such as the hybrid lacZ promoter of the Bluescript® phagemid (Stratagene, LaJolla, CA) or pSport1™ plasmid (Gibco BRL) and the like may be used. The baculovirus polyhedrin promoter may be used in insect cells. Promoters or enhancers derived from the genomes of plant cells (e.g., heat shock, RUBISCO; and storage protein genes) or from plant viruses (e.g., viral promoters or leader sequences) may be cloned into the vector. In mammalian cell systems, promoters from mammalian genes or from mammalian viruses are preferable. If it is necessary to generate a cell line that contains multiple copies of the sequence encoding GRODIF, vectors based on SV40 or EBV may be used with an appropriate selectable marker.

In bacterial systems, a number of expression vectors may be selected depending upon the use intended for GRODIF. For example, when large quantities of GRODIF are needed for the induction of antibodies, vectors which direct high level expression of fusion proteins that are readily purified may be used. Such vectors include, but are not limited to, the multifunctional E. coli cloning and expression vectors such as Bluescript® (Stratagene), in which the sequence encoding GRODIF may be ligated into the vector in frame with sequences for the amino-terminal Met and the subsequent 7 residues of ß-galactosidase so that a hybrid protein is produced; pIN vectors (Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509); and the like. pGEX vectors (Promega, Madison, WI) may also be used to express foreign polypeptides as fusion proteins with glutathione S-transferase (GST). In general, such fusion proteins are soluble and can easily be purified from lysed cells by adsorption to glutathione-agarose beads followed by elution in the presence of free glutathione. Proteins made in such systems may be designed to include heparin, thrombin, or factor XA protease cleavage sites so that the cloned polypeptide of interest can be released from the GST moiety at will.

In the yeast, Saccharomyces cerevisiae, a number of vectors containing constitutive or inducible promoters such as alpha factor, alcohol oxidase, and PGH may be used. For reviews, see Ausubel et al. (supra) and Grant et al. (1987) Methods Enzymol. 153:516-544.

In cases where plant expression vectors are used, the expression of sequences encoding GRODIF may be driven by any of a number of promoters. For example, viral promoters such as the 35S and 19S promoters of CaMV may be used alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) EMBO J. 6:307-311). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used (Coruzzi, G. et al. (1984) EMBO J. 3:1671-1680; Broglie, R. et al. (1984) Science 224:838-843; and Winter, J. et al. (1991) Results Probl. Cell Differ. 17:85-105). These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection. Such techniques are described in a number of generally available reviews (see, for example, Hobbs, S. or Murry, L.E. in McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York, NY; pp. 191-196.

An insect system may also be used to express GRODIF. For example, in one such system, Autographa californica nuclear polyhedrosis virus (AcNPV) is used as a vector to express foreign genes in Spodoptera frugiperda cells or in Trichoplusia larvae. The sequences encoding GRODIF may be cloned into a non-essential region of the virus, such as the polyhedrin gene, and placed under control of the polyhedrin promoter. Successful insertion of GRODIF will render the polyhedrin gene inactive and produce recombinant virus lacking coat protein. The recombinant viruses may then be used to infect, for example, S. frugiperda cells or Trichoplusia larvae in which GRODIF may be expressed (Engelhard, E.K. et al. (1994) Proc. Nat. Acad. Sci. 91:3224-3227).

In mammalian host cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, sequences encoding GRODIF may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain a viable virus which is capable of expressing GRODIF in infected host cells (Logan, J. and Shenk, T. (1984) Proc. Natl. Acad. Sci. 81:3655-3659). In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells.

Specific initiation signals may also be used to achieve more efficient translation of sequences encoding GRODIF. Such signals include the ATG initiation codon and adjacent sequences. In cases where sequences encoding GRODIF, its initiation codon, and upstream sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a

portion thereof, is inserted, exogenous translational control signals including the ATG initiation codon should be provided. Furthermore, the initiation codon should be in the correct reading frame to ensure translation of the entire insert. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers which are appropriate for the particular cell system which is used, such as those described in the literature (Scharf, D. et al. (1994) Results Probl. Cell Differ. 20:125-162).

In addition, a host cell strain may be chosen for its ability to modulate the expression of the inserted sequences or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" form of the protein may also be used to facilitate correct insertion, folding and/or function. Different host cells such as CHO, HeLa, MDCK, HEK293, and WI38, which have specific cellular machinery and characteristic mechanisms for such post-translational activities, may be chosen to ensure the correct modification and processing of the foreign protein.

For long-term, high-yield production of recombinant proteins, stable expression is preferred. For example, cell lines which stably express GRODIF may be transformed using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for 1-2 days in an enriched media before they are switched to selective media. The purpose of the selectable marker is to confer resistance to selection, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be proliferated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase (Wigler, M. et al. (1977) Cell 11:223-32) and adenine phosphoribosyltransferase (Lowy, I. et al. (1980) Cell 22:817-23) genes which can be employed in tk^- or $aprt^-$ cells, respectively. Also, antimetabolite, antibiotic or herbicide resistance can be used as the basis for selection; for example, dhfr which confers resistance to methotrexate (Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. 77:3567-70); npt, which confers resistance to the aminoglycosides neomycin and G-418 (Colbere-Garapin, F. et al (1981) J. Mol. Biol. 150:1-14) and als or pat, which confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively (Murry, supra). Additional selectable genes have

been described, for example, trpB, which allows cells to utilize indole in place of tryptophan, or hisD, which allows cells to utilize histinol in place of histidine (Hartman, S.C. and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. 85:8047-51). Recently, the use of visible markers has gained popularity with such markers as anthocyanins, ß glucuronidase and its substrate GUS, and luciferase and its substrate luciferin, being widely used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system (Rhodes, C.A. et al. (1995) Methods Mol. Biol. 55:121-131).

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, its presence and expression may need to be confirmed. For example, if the sequence encoding GRODIF is inserted within a marker gene sequence, recombinant cells containing sequences encoding GRODIF can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding GRODIF under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

Alternatively, host cells which contain the nucleic acid sequence encoding GRODIF and express GRODIF may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein.

The presence of polynucleotide sequences encoding GRODIF can be detected by DNA-DNA or DNA-RNA hybridization or amplification using probes or portions or fragments of polynucleotides encoding GRODIF. Nucleic acid amplification based assays involve the use of oligonucleotides or oligomers based on the sequences encoding GRODIF to detect transformants containing DNA or RNA encoding GRODIF. As used herein "oligonucleotides" or "oligomers" refer to a nucleic acid sequence of at least about 10 nucleotides and as many as about 60 nucleotides, preferably about 15 to 30 nucleotides, and more preferably about 20-25 nucleotides, which can be used as a probe or amplimer.

A variety of protocols for detecting and measuring the expression of GRODIF, using either polyclonal or monoclonal antibodies specific for the protein are known in the art. Examples include enzyme-linked immunosorbent assay (ELISA), radioimmunoassay (RIA), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on GRODIF is preferred, but a competitive binding assay may be employed. These and other assays are described, among other

places, in Hampton, R. et al. (1990; Serological Methods, a Laboratory Manual, APS Press, St Paul, MN) and Maddox, D.E. et al. (1983; J. Exp. Med. 158:1211-1216).

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding GRODIF include oligolabeling, nick translation, end-labeling or PCR amplification using a labeled nucleotide. Alternatively, the sequences encoding GRODIF, or any portions thereof may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes in vitro by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits (Pharmacia & Upjohn, (Kalamazoo, MI); Promega (Madison WI); and U.S. Biochemical Corp., Cleveland, OH). Suitable reporter molecules or labels, which may be used, include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with nucleotide sequences encoding GRODIF may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a recombinant cell may be secreted or contained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode GRODIF may be designed to contain signal sequences which direct secretion of GRODIF through a prokaryotic or eukaryotic cell membrane. Other recombinant constructions may be used to join sequences encoding GRODIF to nucleotide sequence encoding a polypeptide domain which will facilitate purification of soluble proteins. Such purification facilitating domains include, but are not limited to, metal chelating peptides such as histidine-tryptophan modules that allow purification on immobilized metals, protein A domains that allow purification on immobilized immunoglobulin, and the domain utilized in the FLAGS extension/affinity purification system (Immunex Corp., Seattle, WA). The inclusion of cleavable linker sequences such as those specific for Factor XA or enterokinase (Invitrogen, San Diego, CA) between the purification domain and GRODIF may be used to facilitate purification. One such expression vector provides for expression of a fusion protein containing GRODIF and a nucleic acid encoding 6 histidine residues preceding a thioredoxin or an enterokinase cleavage site. The histidine residues facilitate purification on IMIAC (immobilized metal ion affinity chromatography as described in Porath, J. et al. (1992, Prot. Exp. Purif. 3:263-281) while the

enterokinase cleavage site provides a means for purifying GRODIF from the fusion protein. A discussion of vectors which contain fusion proteins is provided in Kroll, D.J. et al. (1993; DNA Cell Biol. 12:441-453).

In addition to recombinant production, fragments of GRODIF may be produced by direct peptide synthesis using solid-phase techniques Merrifield J. (1963) J. Am. Chem. Soc. 85:2149-2154). Protein synthesis may be performed using manual techniques or by automation. Automated synthesis may be achieved, for example, using Applied Biosystems 431A Peptide Synthesizer (Perkin Elmer). Various fragments of GRODIF may be chemically synthesized separately and combined using chemical methods to produce the full length molecule.

THERAPEUTICS

Chemical and structural homology exists among GRODIF (266285; SEQ ID NO:1), human GDF protein (WO94/15965; SEQ ID NO:3), and GDF-3 from mouse (GI 567205; SEQ ID NO:4). In addition, GRODIF is expressed in testicular tumor and teratocarcinoma cells and murine GDF-3 in immune and hematopoietic tissues. Therefore, GRODIF, appears to have a role in cellular proliferation.

In one embodiment, GRODIF or a fragment or derivative thereof may be added to a cell to stimulate cell proliferation. In particular, GRODIF may be used to encourage wound healing, tissue regeneration or repair in culture, ex vivo, or in vivo. Tissues especially responsive to GRODIF would include connective tissues following tissue destruction by burning or other trauma, spermatogenic tissues to alleviate male sterility, and hemopoietic cells when they are suppressed during or following chemotherapy or by disorders such as AIDS or HIV.

In another embodiment, a vector capable of expressing GRODIF, or a fragment or a derivative thereof, may also be administered to a subject to stimulate cell proliferation as described above.

In another embodiment, an antagonist of GRODIF may be administered to a subject to treat or prevent disorders which are associated with abnormal cell proliferation. In particular, these disorders include cancers such as adenocarcinoma, leukemia, lymphoma, melanoma or sarcoma; immunological disorders including, but not limited to, inflammatory and allergic conditions such as allergic rhinitis, osteoarthritis, gout, atopic dermatitis; autoimmune disorders such as asthma, glomerulonephritis, hypereosinophilia, Sjogren's syndrome, scleroderma, hyperthyroidism (Grave's disease), systemic lupus, myasthenia gravis, rheumatoid arthritis, diabetes mellitus, and Crohn's disease; and viral, bacterial, fungal, parasitic, protozoal, and helminthic infections. In one aspect, antibodies which are specific for GRODIF may be used

directly as an antagonist, or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissue which express GRODIF or in other situations where the modulation of GRODIF levels are therapeutically desirable.

In another embodiment, a vector expressing the complementary sequence of the polynucleotide encoding GRODIF may be administered to a subject to treat or prevent disorders associated with cell proliferation including those described above.

In other embodiments, any of the therapeutic proteins, antagonists, antibodies, agonists, antisense sequences or vectors described above may be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

Antagonists or inhibitors of GRODIF may be produced using methods which are generally known in the art. In particular, purified GRODIF may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind GRODIF.

The antibodies may be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, single chain, Fab fragments, and fragments produced by a Fab expression library. Neutralizing antibodies, (i.e., those which inhibit dimer formation) are especially preferred for therapeutic use.

For the production of antibodies, various hosts including goats, rabbits, rats, mice, humans, and others, may be immunized by injection with GRODIF or any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, keyhole limpet hemocyanin, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and Corynebacterium parvum are especially preferable.

It is preferred that the peptides, fragments, or oligopeptides used to induce antibodies to GRODIF have an amino acid sequence consisting of at least five amino acids, and more preferably at least 10 amino acids. It is also preferable that they are identical to a portion of the amino acid sequence of the natural protein, and they may contain the entire amino acid sequence

of a small, naturally occurring molecule. Short stretches of GRODIF amino acids may be fused with those of another protein such as keyhole limpet hemocyanin and antibody produced against the chimeric molecule.

Monoclonal antibodies to GRODIF may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique (Kohler, G. et al. (1975) Nature 256:495-497; Kozbor, D. et al. (1985) J. Immunol. Methods 81:31-42; Cote, R.J. et al. (1983) Proc. Natl. Acad. Sci. 80:2026-2030; Cole, S.P. et al. (1984) Mol. Cell Biol. 62:109-120).

In addition, techniques developed for the production of "chimeric antibodies", the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity can be used (Morrison, S.L. et al. (1984) Proc. Natl. Acad. Sci. 81:6851-6855; Neuberger, M.S. et al. (1984) Nature 312:604-608; Takeda, S. et al. (1985) Nature 314:452-454). Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce GRODIF-specific single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobin libraries (Burton D.R. (1991) Proc. Natl. Acad. Sci. 88:11120-3).

Antibodies may also be produced by inducing in vivo production in the lymphocyte population or by screening recombinant immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature (Orlandi, R. et al. (1989) Proc. Natl. Acad. Sci. 86: 3833-3837; Winter, G. et al. (1991) Nature 349:293-299).

Antibody fragments which contain specific binding sites for GRODIF may also be generated. For example, such fragments include, but are not limited to, the F(ab')2 fragments which can be produced by pepsin digestion of the antibody molecule and the Fab fragments which can be generated by reducing the disulfide bridges of the F(ab')2 fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity (Huse, W.D. et al. (1989) Science 254:1275-1281).

Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between GRODIF and its specific antibody. A two-site, monoclonal-based immunoassay utilizing

monoclonal antibodies reactive to two non-interfering GRODIF epitopes is preferred, but a competitive binding assay may also be employed (Maddox, supra).

In another embodiment of the invention, the polynucleotides encoding GRODIF, or any fragment thereof, or antisense molecules, may be used for therapeutic purposes. In one aspect, antisense to the polynucleotide encoding GRODIF may be used in situations in which it would be desirable to block the transcription of the mRNA. In particular, cells may be transformed with sequences complementary to polynucleotides encoding GRODIF. Thus, antisense molecules may be used to modulate GRODIF activity, or to achieve regulation of gene function. Such technology is now well known in the art, and sense or antisense oligomers or larger fragments, can be designed from various locations along the coding or control regions of sequences encoding GRODIF.

Expression vectors derived from retro viruses, adenovirus, herpes or vaccinia viruses, or from various bacterial plasmids may be used for delivery of nucleotide sequences to the targeted organ, tissue or cell population. Methods which are well known to those skilled in the art can be used to construct recombinant vectors which will express antisense molecules complementary to the polynucleotides of the gene encoding GRODIF. These techniques are described both in Sambrook et al. (supra) and in Ausubel et al. (supra).

Genes encoding GRODIF can be turned off by transforming a cell or tissue with expression vectors which express high levels of a polynucleotide or fragment thereof which encodes GRODIF. Such constructs may be used to introduce untranslatable sense or antisense sequences into a cell. Even in the absence of integration into the DNA, such vectors may continue to transcribe RNA molecules until they are disabled by endogenous nucleases. Transient expression may last for a month or more with a non-replicating vector and even longer if appropriate replication elements are part of the vector system.

As mentioned above, modifications of gene expression can be obtained by designing antisense molecules, DNA, RNA, or PNA, to the control regions of the gene encoding GRODIF, i.e., the promoters, enhancers, and introns. Oligonucleotides derived from the transcription initiation site, e.g., between positions -10 and +10 from the start site, are preferred. Similarly, inhibition can be achieved using "triple helix" base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature (Gee, J.E. et al. (1994) In: Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing Co., Mt.

Kisco, NY). The antisense molecules may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. Examples which may be used include engineered hammerhead motif ribozyme molecules that can specifically and efficiently catalyze endonucleolytic cleavage of sequences encoding GRODIF.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites which include the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides corresponding to the region of the target gene containing the cleavage site may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Antisense molecules and ribozymes of the invention may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by in vitro and in vivo transcription of DNA sequences encoding GRODIF. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize antisense RNA constitutively or inducibly can be introduced into cell lines, cells, or tissues.

RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3' ends of the molecule or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

Many methods for introducing vectors into cells or tissues are available and equally suitable for use in vivo, in vitro, and ex vivo. For ex vivo therapy, vectors may be introduced

into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection and by liposome injections may be achieved using methods which are well known in the art.

Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as dogs, cats, cows, horses, rabbits, monkeys, and most preferably, humans.

An additional embodiment of the invention relates to the administration of a pharmaceutical composition, in conjunction with a pharmaceutically acceptable carrier, for any of the therapeutic effects discussed above. Such pharmaceutical compositions may consist of GRODIF, antibodies to GRODIF, mimetics, agonists, antagonists, or inhibitors of GRODIF. The compositions may be administered alone or in combination with at least one other agent, such as stabilizing compound, which may be administered in any sterile, biocompatible pharmaceutical carrier, including, but not limited to, saline, buffered saline, dextrose, and water. The compositions may be administered to a patient alone, or in combination with other agents, drugs or hormones.

The pharmaceutical compositions utilized in this invention may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

In addition to the active ingredients, these pharmaceutical compositions may contain suitable pharmaceutically-acceptable carriers comprising excipients and auxiliaries which facilitate processing of the active compounds into preparations which can be used pharmaceutically. Further details on techniques for formulation and administration may be found in the latest edition of Remington's Pharmaceutical Sciences (Maack Publishing Co., Easton, PA).

Pharmaceutical compositions for oral administration can be formulated using pharmaceutically acceptable carriers well known in the art in dosages suitable for oral administration. Such carriers enable the pharmaceutical compositions to be formulated as tablets, pills, dragees, capsules, liquids, gels, syrups, slurries, suspensions, and the like, for ingestion by the patient.

Pharmaceutical preparations for oral use can be obtained through combination of active compounds with solid excipient, optionally grinding a resulting mixture, and processing the mixture of granules, after adding suitable auxiliaries, if desired, to obtain tablets or dragee cores.

Suitable excipients are carbohydrate or protein fillers, such as sugars, including lactose, sucrose, mannitol, or sorbitol; starch from corn, wheat, rice, potato, or other plants; cellulose, such as methyl cellulose, hydroxypropylmethyl-cellulose, or sodium carboxymethylcellulose; gums including arabic and tragacanth; and proteins such as gelatin and collagen. If desired, disintegrating or solubilizing agents may be added, such as the cross-linked polyvinyl pyrrolidone, agar, alginic acid, or a salt thereof, such as sodium alginate.

Dragee cores may be used in conjunction with suitable coatings, such as concentrated sugar solutions, which may also contain gum arabic, talc, polyvinylpyrrolidone, carbopol gel, polyethylene glycol, and/or titanium dioxide, lacquer solutions, and suitable organic solvents or solvent mixtures. Dyestuffs or pigments may be added to the tablets or dragee coatings for product identification or to characterize the quantity of active compound, i.e., dosage.

Pharmaceutical preparations which can be used orally include push-fit capsules made of gelatin, as well as soft, sealed capsules made of gelatin and a coating, such as glycerol or sorbitol. Push-fit capsules can contain active ingredients mixed with a filler or binders, such as lactose or starches, lubricants, such as talc or magnesium stearate, and, optionally, stabilizers. In soft capsules, the active compounds may be dissolved or suspended in suitable liquids, such as fatty oils, liquid, or liquid polyethylene glycol with or without stabilizers.

Pharmaceutical formulations suitable for parenteral administration may be formulated in aqueous solutions, preferably in physiologically compatible buffers such as Hanks's solution, Ringer's solution, or physiologically buffered saline. Aqueous injection suspensions may contain substances which increase the viscosity of the suspension, such as sodium carboxymethyl cellulose, sorbitol, or dextran. Additionally, suspensions of the active compounds may be prepared as appropriate oily injection suspensions. Suitable lipophilic solvents or vehicles include fatty oils such as sesame oil, or synthetic fatty acid esters, such as ethyl oleate or triglycerides, or liposomes. Optionally, the suspension may also contain suitable stabilizers or agents which increase the solubility of the compounds to allow for the preparation of highly concentrated solutions.

For topical or nasal administration, penetrants appropriate to the particular barrier to be permeated are used in the formulation. Such penetrants are generally known in the art.

The pharmaceutical compositions of the present invention may be manufactured in a manner that is known in the art, e.g., by means of conventional mixing, dissolving, granulating, dragee-making, levigating, emulsifying, encapsulating, entrapping, or lyophilizing processes.

The pharmaceutical composition may be provided as a salt and can be formed with many

acids, including but not limited to, hydrochloric, sulfuric, acetic, lactic, tartaric, malic, succinic, etc. Salts tend to be more soluble in aqueous or other protonic solvents than are the corresponding free base forms. In other cases, the preferred preparation may be a lyophilized powder which may contain any or all of the following: 1-50 mM histidine, 0.1%-2% sucrose, and 2-7% mannitol, at a pH range of 4.5 to 5.5, that is combined with buffer prior to use.

After pharmaceutical compositions have been prepared, they can be placed in an appropriate container and labeled for treatment of an indicated condition. For administration of GRODIF, such labeling would include amount, frequency, and method of administration.

Pharmaceutical compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells, or in animal models, usually mice, rabbits, dogs, or pigs. The animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example GRODIF or fragments thereof, antibodies of GRODIF, agonists, antagonists or inhibitors of GRODIF, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or experimental animals, e.g., ED50 (the dose therapeutically effective in 50% of the population) and LD50 (the dose lethal to 50% of the population). The dose ratio between therapeutic and toxic effects is the therapeutic index, and it can be expressed as the ratio, LD50/ED50. Pharmaceutical compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies is used in formulating a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that include the ED50 with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the subject that requires treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, general health of the subject, age, weight, and

gender of the subject, diet, time and frequency of administration, drug combination(s), reaction sensitivities, and tolerance/response to therapy. Long-acting pharmaceutical compositions may be administered every 3 to 4 days, every week, or once every two weeks depending on half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from 0.1 to 100,000 micrograms, up to a total dose of about 1 g, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

DIAGNOSTICS

In another embodiment, antibodies which specifically bind GRODIF may be used for the diagnosis of conditions or diseases characterized by expression of GRODIF, or in assays to monitor patients being treated with GRODIF, agonists, antagonists or inhibitors. The antibodies useful for diagnostic purposes may be prepared in the same manner as those described above for therapeutics. Diagnostic assays for GRODIF include methods which utilize the antibody and a label to detect GRODIF in human body fluids or extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by joining them, either covalently or non-covalently, with a reporter molecule. A wide variety of reporter molecules which are known in the art may be used, several of which are described above.

A variety of protocols including ELISA, RIA, and FACS for measuring GRODIF are known in the art and provide a basis for diagnosing altered or abnormal levels of GRODIF expression. Normal or standard values for GRODIF expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, preferably human, with antibody to GRODIF under conditions suitable for complex formation. The amount of standard complex formation may be quantified by various methods, but preferably by photometric, means. Quantities of GRODIF expressed in subject, control and disease, samples from biopsied tissues are compared with the standard values. Deviation between standard and subject values establishes the parameters for diagnosing disease.

In another embodiment of the invention, the polynucleotides encoding GRODIF may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotide sequences, antisense RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantitate gene expression in biopsied tissues in which expression of GRODIF may be

correlated with disease. The diagnostic assay may be used to distinguish between absence, presence, and excess expression of GRODIF, and to monitor regulation of GRODIF levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotide sequences, including genomic sequences, encoding GRODIF or closely related molecules, may be used to identify nucleic acid sequences which encode GRODIF. The specificity of the probe, whether it is made from a highly specific region, e.g., 10 unique nucleotides in the 5' regulatory region, or a less specific region, e.g., especially in the 3' coding region, and the stringency of the hybridization or amplification (maximal, high, intermediate, or low) will determine whether the probe identifies only naturally occurring sequences encoding GRODIF, alleles, or related sequences.

Probes may also be used for the detection of related sequences, and should preferably contain at least 50% of the nucleotides from any of the GRODIF encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and derived from the nucleotide sequence of SEQ ID NO:2 or from genomic sequence including promoter, enhancer elements, and introns of the naturally occurring GRODIF.

Means for producing specific hybridization probes for DNAs encoding GRODIF include the cloning of nucleic acid sequences encoding GRODIF or GRODIF derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, commercially available, and may be used to synthesize RNA probes in vitro by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter groups, for example, radionuclides such as 32P or 35S, or enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotide sequences encoding GRODIF may be used for the diagnosis of disorders which are associated with expression of GRODIF. Examples of such disorders include cancers such as adenocarcinoma, leukemia, lymphoma, melanoma or sarcoma; and immunological disorders such as allergic rhinitis, osteoarthritis, gout, atopic dermatitis; autoimmune disorders such as asthma, glomerulonephritis, hypereosinophilia, Sjogren's syndrome, scleroderma, hyperthyroidism (Grave's disease), systemic lupus, myasthenia gravis, rheumatoid arthritis, diabetes mellitus, and Crohn's disease; and viral, bacterial, fungal, parasitic, protozoal, and helminthic infections.

The polynucleotide sequences encoding GRODIF may be used in Southern or northern

analysis, dot blot, or other membrane-based technologies; in PCR technologies; or in dip stick, pin, ELISA or chip assays utilizing fluids or tissues from patient biopsies to detect altered GRODIF expression. Such qualitative or quantitative methods are well known in the art.

In a particular aspect, the nucleotide sequences encoding GRODIF may be useful in assays that detect activation or induction of various cancers, particularly those mentioned above. The nucleotide sequences encoding GRODIF may be labeled by standard methods, and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantitated and compared with a standard value. If the amount of signal in the biopsied or extracted sample is significantly altered from that of a comparable control sample, the nucleotide sequences have hybridized with nucleotide sequences in the sample, and the presence of altered levels of nucleotide sequences encoding GRODIF in the sample indicates the presence of the associated disease. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or in monitoring the treatment of an individual patient.

In order to provide a basis for the diagnosis of disease associated with expression of GRODIF, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, which encodes GRODIF, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with those from an experiment where a known amount of a substantially purified polynucleotide is used. Standard values obtained from normal samples may be compared with values obtained from samples from patients who are symptomatic for disease. Deviation between standard and subject values is used to establish the presence of disease.

Once disease is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to evaluate whether the level of expression in the patient begins to approximate that which is observed in the normal patient. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

With respect to cancer, the presence of a relatively high amount of transcript in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ

preventative measures or aggressive treatment earlier thereby preventing the development or further progression of the cancer.

Additional diagnostic uses for oligonucleotides designed from the sequences encoding GRODIF may involve the use of PCR. Such oligomers may be chemically synthesized, generated enzymatically, or produced from a recombinant source. Oligomers will preferably consist of two nucleotide sequences, one with sense orientation (5'->3') and another with antisense (3'<-5'), employed under optimized conditions for identification of a specific gene or condition. The same two oligomers, nested sets of oligomers, or even a degenerate pool of oligomers may be employed under less stringent conditions for detection and/or quantitation of closely related DNA or RNA sequences.

Methods which may also be used to quantitate the expression of GRODIF include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and standard curves onto which the experimental results are interpolated (Melby, P.C. et al. (1993) J. Immunol. Methods, 159:235-244; Duplaa, C. et al. (1993) Anal. Biochem. 229-236). The speed of quantitation of multiple samples may be accelerated by running the assay in an ELISA format where the oligomer of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

In another embodiment of the invention, the nucleic acid sequences which encode GRODIF may also be used to generate hybridization probes which are useful for mapping the naturally occurring genomic sequence. The sequences may be mapped to a particular chromosome or to a specific region of the chromosome using well known techniques. Such techniques include FISH, FACS, or artificial chromosome constructions, such as yeast artificial chromosomes, bacterial artificial chromosomes, bacterial P1 constructions or single chromosome cDNA libraries as reviewed in Price, C.M. (1993) Blood Rev. 7:127-134, and Trask, B.J. (1991) Trends Genet. 7:149-154.

FISH (as described in Verma et al. (1988) Human Chromosomes: A Manual of Basic Techniques, Pergamon Press, New York, NY) may be correlated with other physical chromosome mapping techniques and genetic map data. Examples of genetic map data can be found in the 1994 Genome Issue of Science (265:1981f). Correlation between the location of the gene encoding GRODIF on a physical chromosomal map and a specific disease, or predisposition to a specific disease, may help delimit the region of DNA associated with that genetic disease. The nucleotide sequences of the subject invention may be used to detect differences in gene sequences between normal, carrier, or affected individuals.

In situ hybridization of chromosomal preparations and physical mapping techniques such as linkage analysis using established chromosomal markers may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the number or arm of a particular human chromosome is not known. New sequences can be assigned to chromosomal arms, or parts thereof, by physical mapping. This provides valuable information to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the disease or syndrome has been crudely localized by genetic linkage to a particular genomic region, for example, AT to 11q22-23 (Gatti, R.A. et al. (1988) Nature 336:577-580), any sequences mapping to that area may represent associated or regulatory genes for further investigation. The nucleotide sequence of the subject invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc. among normal, carrier, or affected individuals.

In another embodiment of the invention, GRODIF, its catalytic or immunogenic fragments or oligopeptides thereof, can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes, between GRODIF and the agent being tested, may be measured.

Another technique for drug screening which may be used provides for high throughput screening of compounds having suitable binding affinity to the protein of interest as described in published PCT application WO84/03564. In this method, as applied to GRODIF large numbers of different small test compounds are synthesized on a solid substrate, such as plastic pins or some other surface. The test compounds are reacted with GRODIF, or fragments thereof, and washed. Bound GRODIF is then detected by methods well known in the art. Purified GRODIF can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding GRODIF specifically compete with a test compound for binding GRODIF. In this manner, the antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with GRODIF.

In additional embodiments, the nucleotide sequences which encode GRODIF may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to,

such properties as the triplet genetic code and specific base pair interactions.

The examples below are provided to illustrate the subject invention and are not included for the purpose of limiting the invention.

INDUSTRIAL APPLICABILITY

I HNT2NOT01 cDNA Library Construction

The hNT2 cell line exhibits characteristics of a committed neuronal precursor cell which is still at an early stage of development. The HNT2NOT01 library was constructed by Stratagene. The mRNA was isolated, and cDNAs were primed using oligo d(T) technology. Synthetic adapter oligonucleotides were ligated onto the cDNA molecules which were 500 BP (or larger), and these molecules were inserted into the Uni-ZAP™ vector system (Stratagene).

The quality of the cDNA library was screened using DNA probes, and the pBluescript™ phagemid (Stratagene) was excised. The custom-constructed library phage particles were infected into E. coli host strain XL1-Blue™ (Stratagene). Alternative unidirectional vectors include, but are not limited to, pcDNAI (Invitrogen, San Diego CA) and pSHlox-1 (Novagen, Madison WI).

II Isolation and Sequencing of cDNA Clones

The phagemid forms of individual cDNA clones were obtained by employing the Miniprep Kit (Catalog No. 77468) available from Advanced Genetic Technologies Corp., Gaithersburg MD. This kit is in the 96-well format and provides enough reagents for 960 purifications. Each kit is provided with a recommended protocol which was employed except for the following changes. First, the 96 wells were each filled with only 1 ml of sterile terrific broth with carbenicillin at 25 mg/L and glycerol at 0.4%. After the wells were inoculated, the bacteria were cultured for 24 hours and lysed with 60 µl of lysis buffer. A centrifugation step (2900 rpm for 5 minutes) was performed before the contents of the block were added to the primary filter plate. The optional step of adding isopropanol to TRIS buffer was not routinely performed. After the last step in the protocol, samples were transferred to a Beckman 96-well block for storage.

Alternatively, the in vivo excision process, in which the host bacterial strain is co-infected with both the library phage and an f1 helper phage. Polypeptides or enzymes derived from both the library-containing phage and the helper phage nick the DNA, initiate new DNA synthesis from defined sequences on the target DNA, and create a smaller, single stranded circular phagemid DNA molecule that includes all DNA sequences of the pBluescript phagemid and the cDNA insert. The phagemid DNA is released from the cells, purified, and used to reinfect fresh

host cells (SOLR, Stratagene) where double-stranded phagemid DNA is produced. Because the phagemid carries the gene for β-lactamase, the newly transformed bacteria are selected on medium containing ampicillin.

Phagemid DNA may also be purified using the QIAWELL-8 Plasmid purification system (QIAGEN Inc, Chatsworth CA). The DNA was eluted from the purification resin and prepared for DNA sequencing and other analytical manipulations.

III Homology Searching of cDNA Clones and Their Deduced Proteins

Each cDNA was compared to sequences in GenBank using a search algorithm developed by Applied Biosystems and incorporated into the INHERIT™ 670 sequence analysis system. In this algorithm, Pattern Specification Language (TRW Inc, Los Angeles, CA) was used to determine regions of homology. The three parameters that determine how the sequence comparisons run were window size, window offset, and error tolerance. Using a combination of these three parameters, the DNA database was searched for sequences containing regions of homology to the query sequence, and the appropriate sequences were scored with an initial value. Subsequently, these homologous regions were examined using dot matrix homology plots to distinguish regions of homology from chance matches. Smith-Waterman alignments were used to display the results of the homology search.

Peptide and protein sequence homologies were ascertained using the INHERIT- 670 sequence analysis system using the methods similar to those used in DNA sequence homologies. Pattern Specification Language and parameter windows were used to search protein databases for sequences containing regions of homology which were scored with an initial value. Dot-matrix homology plots were examined to distinguish regions of significant homology from chance matches.

BLAST, which stands for Basic Local Alignment Search Tool (Altschul, S.F. (1993) J. Mol. Evol. 36:290-300; Altschul et al. (1990) J. Mol. Biol. 215:403-410), was used to search for local sequence alignments. BLAST produces alignments of both nucleotide and amino acid sequences to determine sequence similarity. Because of the local nature of the alignments, BLAST is especially useful in determining exact matches or in identifying homologs. BLAST is useful for matches which do not contain gaps. The fundamental unit of BLAST algorithm output is the High-scoring Segment Pair (HSP).

An HSP consists of two sequence fragments of arbitrary but equal lengths whose alignment is locally maximal and for which the alignment score meets or exceeds a threshold or cutoff score set by the user. The BLAST approach is to look for HSPs between a query sequence

and a database sequence, to evaluate the statistical significance of any matches found, and to report only those matches which satisfy the user-selected threshold of significance. The parameter E establishes the statistically significant threshold for reporting database sequence matches. E is interpreted as the upper bound of the expected frequency of chance occurrence of an HSP (or set of HSPs) within the context of the entire database search. Any database sequence whose match satisfies E is reported in the program output.

IV Northern Analysis

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound (Sambrook et al., supra).

Analogous computer techniques using BLAST (Altschul, S.F. 1993 and 1990, supra) are used to search for identical or related molecules in nucleotide databases such as GenBank or the LIFESEQ™ database (Incyte Pharmaceuticals). This analysis is much faster than multiple, membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or homologous.

The basis of the search is the product score which is defined as:

$$\frac{\%\ \text{sequence identity} \times \%\ \text{maximum BLAST score}}{100}$$

The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. For example, with a product score of 40, the match will be exact within a 1-2% error; and at 70, the match will be exact. Homologous molecules are usually identified by selecting those which show product scores between 15 and 40, although lower scores may identify related molecules.

The results of northern analysis are reported as a list of libraries in which the transcript encoding GRODIF occurs. Abundance and percent abundance are also reported. Abundance directly reflects the number of times a particular transcript is represented in a cDNA library, and percent abundance is abundance divided by the total number of sequences examined in the cDNA library.

V Extension of GRODIF-Encoding Polynucleotides

Nucleic acid sequences of Incyte Clone 266285 or SEQ ID NO:2 are used to design oligonucleotide primers for extending a partial nucleotide sequence to full length or for obtaining 5' or 3', intron or other control sequences from genomic libraries. One primer is synthesized to initiate extension in the antisense direction (XLR) and the other is synthesized to extend sequence

in the sense direction (XLF). Primers are used to facilitate the extension of the known sequence "outward" generating amplicons containing new, unknown nucleotide sequence for the region of interest. The initial primers are designed from the cDNA using OLIGO 4.06 (National Biosciences), or another appropriate program, to be 22-30 nucleotides in length, to have a GC content of 50% or more, and to anneal to the target sequence at temperatures about 68°-72° C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations is avoided.

The original, selected cDNA libraries, or a human genomic library are used to extend the sequence; the latter is most useful to obtain 5' upstream regions. If more extension is necessary or desired; additional sets of primers are designed to further extend the known region.

By following the instructions for the XL-PCR kit (Perkin Elmer) and thoroughly mixing the enzyme and reaction mix, high fidelity amplification is obtained. Beginning with 40 pmol of each primer and the recommended concentrations of all other components of the kit, PCR is performed using the Peltier Thermal Cycler (PTC200; M.J. Research, Watertown, MA) and the following parameters:

Step 1	94° C for 1 min (initial denaturation)
Step 2	65° C for 1 min
Step 3	68° C for 6 min
Step 4	94° C for 15 sec
Step 5	65° C for 1 min
Step 6	68° C for 7 min
Step 7	Repeat step 4-6 for 15 additional cycles
Step 8	94° C for 15 sec
Step 9	65° C for 1 min
Step 10	68° C for 7:15 min
Step 11	Repeat step 8-10 for 12 cycles
Step 12	72° C for 8 min
Step 13	4° C (and holding)

A 5-10 μl aliquot of the reaction mixture is analyzed by electrophoresis on a low concentration (about 0.6-0.8%) agarose mini-gel to determine which reactions were successful in extending the sequence. Bands thought to contain the largest products are selected and removed from the gel. Further purification involves using a commercial gel extraction method such as QIAQuick™ (QIAGEN Inc., Chatsworth, CA). After recovery of the DNA, Klenow enzyme is used to trim single-stranded, nucleotide overhangs creating blunt ends which facilitate religation and cloning.

After ethanol precipitation, the products are redissolved in 13 μl of ligation buffer, 1μl

T4-DNA ligase (15 units) and 1μl T4 polynucleotide kinase are added, and the mixture is incubated at room temperature for 2-3 hours or overnight at 16° C. Competent E. coli cells (in 40 μl of appropriate media) are transformed with 3 μl of ligation mixture and cultured in 80 μl of SOC medium (Sambrook et al., supra). After incubation for one hour at 37° C, the whole transformation mixture is plated on Luria Bertani (LB)-agar (Sambrook et al., supra) containing 2x Carb. The following day, several colonies are randomly picked from each plate and cultured in 150 μl of liquid LB/2x Carb medium placed in an individual well of an appropriate, commercially-available, sterile 96-well microtiter plate. The following day, 5 μl of each overnight culture is transferred into a non-sterile 96-well plate and after dilution 1:10 with water, 5 μl of each sample is transferred into a PCR array.

For PCR amplification, 18 μl of concentrated PCR reaction mix (3.3x) containing 4 units of rTth DNA polymerase, a vector primer, and one or both of the gene specific primers used for the extension reaction are added to each well. Amplification is performed using the following conditions:

Step 1	94° C for 60 sec
Step 2	94° C for 20 sec
Step 3	55° C for 30 sec
Step 4	72° C for 90 sec
Step 5	Repeat steps 2-4 for an additional 29 cycles
Step 6	72° C for 180 sec
Step 7	4° C (and holding)

Aliquots of the PCR reactions are run on agarose gels together with molecular weight markers. The sizes of the PCR products are compared to the original partial cDNAs, and appropriate clones are selected, ligated into plasmid, and sequenced.

VI Labeling and Use of Hybridization Probes

Hybridization probes derived from SEQ ID NO:2 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base-pairs, is specifically described, essentially the same procedure is used with larger cDNA fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 (National Biosciences), labeled by combining 50 pmol of each oligomer and 250 μCi of [γ-^{32}P] adenosine triphosphate (Amersham) and T4 polynucleotide kinase (DuPont NEN®, Boston, MA). The labeled oligonucleotides are substantially purified with Sephadex G-25 superfine resin column (Pharmacia & Upjohn). A portion containing 10^7 counts per minute of each of the sense and antisense oligonucleotides is used in a typical membrane based hybridization analysis of

human genomic DNA digested with one of the following endonucleases (Ase I, Bgl II, Eco RI, Pst I, Xba 1, or Pvu II; DuPont NEN®).

The DNA from each digest is fractionated on a 0.7 percent agarose gel and transferred to nylon membranes (Nytran Plus, Schleicher & Schuell, Durham, NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under increasingly stringent conditions up to 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. After XOMAT AR™ film (Kodak, Rochester, NY) is exposed to the blots in a Phosphoimager cassette (Molecular Dynamics, Sunnyvale, CA) for several hours, hybridization patterns are compared visually.

VII Complementary Polynucleotides

Sequence complementary to the GRODIF-encoding sequence, or any part thereof, is used to inhibit in vivo or in vitro expression of naturally occurring GRODIF. Although use of oligonucleotides, comprising about 20 base-pairs, is specifically described, essentially the same procedure is used with larger sequence fragments. An oligonucleotide based on the coding sequence of GRODIF, as shown in Figures 1A, 1B, 1C and 1D, is used to inhibit expression of naturally occurring GRODIF. The complementary oligonucleotide is designed from the most unique 5' sequence and used either to inhibit transcription by preventing promoter binding to the upstream nontranslated sequence or translation of an GRODIF-encoding transcript by preventing ribosomal binding. Using an appropriate portion of the signal or 5' sequence of SEQ ID NO:2, an effective complementary oligonucleotide includes any 15-30 nucleotides spanning the region which translates into the signal or 5' coding sequence of the GRODIF.

VIII Expression of GRODIF

Expression of GRODIF is accomplished by subcloning the cDNAs into appropriate vectors and transforming the vectors into host cells. In this case, the cloning vector, pBluescript™ previously used for the generation of the cDNA library is used to express GRODIF in E. coli. Upstream of the cloning site, this vector contains a promoter for ß-galactosidase, followed by sequence containing the amino-terminal Met, and the subsequent seven residues of ß-galactosidase. Immediately following these eight residues is a bacteriophage promoter useful for transcription and a linker containing a number of unique restriction sites.

Induction of an isolated, transformed bacterial strain with IPTG using standard methods produces a fusion protein which consists of the first eight residues of ß-galactosidase, about 5 to 15 residues of linker, and the full length protein. The signal residues direct the secretion of GRODIF into the bacterial growth media which can be used directly in the following assay for

activity.

IX Demonstration of GRODIF Activity

Freshly isolated murine thymocytes are incubated in the appropriate tissue culture media with phytohemagglutinin (PHA) and ^{3}H- thymidine. Increasing concentrations of GRODIF are then added to the cultures, and they are incubated for 4 hrs at 37°C. PHA stimulation of thymocyte proliferation is assessed by quantitating the incorporation of radionucleotides into DNA by radio-scintillation counting. Addition of increasing concentrations of GRODIF to the cultures produces a proportional stimulation of mitogen-driven thymocyte proliferation compared to untreated controls.

X Production of GRODIF Specific Antibodies

GRODIF that is substantially purified using PAGE electrophoresis (Sambrook, supra), or other purification techniques, is used to immunize rabbits and to produce antibodies using standard protocols. The amino acid sequence deduced from SEQ ID NO:2 is analyzed using DNASTAR software (DNASTAR Inc) to determine regions of high immunogenicity and a corresponding oligopolypeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions, is described by Ausubel et al. (supra), and others.

Typically, the oligopeptides are 15 residues in length, synthesized using an Applied Biosystems Peptide Synthesizer Model 431A using fmoc-chemistry, and coupled to keyhole limpet hemocyanin (KLH, Sigma, St. Louis, MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS; Ausubel et al., supra). Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. The resulting antisera are tested for antipeptide activity, for example, by binding the peptide to plastic, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radioiodinated, goat anti-rabbit IgG.

XI Purification of Naturally Occurring GRODIF Using Specific Antibodies

Naturally occurring or recombinant GRODIF is substantially purified by immunoaffinity chromatography using antibodies specific for GRODIF. An immunoaffinity column is constructed by covalently coupling GRODIF antibody to an activated chromatographic resin, such as CnBr-activated Sepharose (Pharmacia & Upjohn). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

Media containing GRODIF is passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of GRODIF (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt

antibody/GRODIF binding (eg, a buffer of pH 2-3 or a high concentration of a chaotrope, such as urea or thiocyanate ion), and GRODIF is collected.

XII Identification of Molecules Which Interact with GRODIF

GRODIF or biologically active fragments thereof are labeled with ^{125}I Bolton-Hunter reagent (Bolton et al. (1973) Biochem. J. 133:529). Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled GRODIF, washed and any wells with labeled GRODIF complex are assayed. Data obtained using different concentrations of GRODIF are used to calculate values for the number, affinity, and association of GRODIF with the candidate molecules.

All publications and patents mentioned in the above specification are herein incorporated by reference. Various modifications and variations of the described method and system of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with specific preferred embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are obvious to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

SEQUENCE LISTING

(1) GENERAL INFORMATION

(i) APPLICANT: INCYTE PHARMACEUTICALS, INC.

(ii) TITLE OF THE INVENTION: NOVEL HUMAN GROWTH AND DIFFERENTIATION PROTEIN

(iii) NUMBER OF SEQUENCES: 4

(iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: Incyte Pharmaceuticals, Inc.
(B) STREET: 3174 Porter Drive
(C) CITY: Palo Alto
(D) STATE: CA
(E) COUNTRY: USA
(F) ZIP: 94304

(v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Diskette
(B) COMPUTER: IBM Compatible
(C) OPERATING SYSTEM: DOS
(D) SOFTWARE: FastSEQ for Windows Version 2.0

(vi) CURRENT APPLICATION DATA:
(A) PCT APPLICATION NUMBER: To Be Assigned
(B) FILING DATE: Herewith
(C) CLASSIFICATION:

(vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: US 08/822,259
(B) FILING DATE: 20-MAR-1997

(viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: Billings, Lucy J.
(B) REGISTRATION NUMBER: 36,749
(C) REFERENCE/DOCKET NUMBER: PF-0231 PCT

(ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: 650-855-0555
(B) TELEFAX: 650-845-4166
(C) TELEX:

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 364 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:
(A) LIBRARY: HNT2NOT01
(B) CLONE: 266285

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

```
Met Leu Arg Phe Leu Pro Asp Leu Ala Phe Ser Phe Leu Leu Ile Leu
 1               5                  10                  15
Ala Leu Gly Gln Ala Val Gln Phe Gln Glu Tyr Val Phe Leu Gln Phe
            20                  25                  30
Leu Gly Leu Asp Lys Ala Pro Ser Pro Gln Lys Phe Gln Pro Val Pro
        35                  40                  45
Tyr Ile Leu Lys Lys Ile Phe Gln Asp Arg Glu Ala Ala Ala Thr Thr
    50                  55                  60
Gly Val Ser Arg Asp Leu Cys Tyr Val Lys Glu Leu Gly Val Arg Gly
65                  70                  75                  80
Asn Val Leu Arg Phe Leu Pro Asp Gln Gly Phe Phe Leu Tyr Pro Lys
                85                  90                  95
Lys Ile Ser Gln Ala Ser Ser Cys Leu Gln Lys Leu Leu Tyr Phe Asn
            100                 105                 110
Leu Ser Ala Ile Lys Glu Arg Glu Gln Leu Thr Leu Ala Gln Leu Gly
        115                 120                 125
Leu Asp Leu Gly Pro Asn Ser Tyr Tyr Asn Leu Gly Pro Glu Leu Glu
    130                 135                 140
Leu Ala Leu Phe Leu Val Gln Glu Pro His Val Trp Gly Gln Thr Thr
145                 150                 155                 160
Pro Lys Pro Gly Lys Met Phe Val Leu Arg Ser Val Pro Trp Pro Gln
                165                 170                 175
Gly Ala Val His Phe Asn Leu Leu Asp Val Ala Lys Asp Trp Asn Asp
            180                 185                 190
Asn Pro Arg Lys Asn Phe Gly Leu Phe Leu Glu Ile Leu Val Lys Glu
        195                 200                 205
Asp Arg Asp Ser Gly Val Asn Phe Gln Pro Glu Asp Asn Cys Ala Arg
    210                 215                 220
Leu Arg Cys Ser Leu His Ala Ser Leu Leu Val Val Thr Leu Asn Pro
225                 230                 235                 240
Asp Gln Cys His Pro Ser Arg Lys Arg Arg Ala Ala Ile Pro Val Pro
                245                 250                 255
Lys Leu Ser Cys Lys Asn Leu Cys His Arg His Gln Leu Phe Ile Asn
            260                 265                 270
Phe Arg Asp Leu Gly Trp His Lys Trp Ile Ile Ala Pro Lys Gly Phe
        275                 280                 285
Met Ala Asn Tyr Cys His Gly Glu Cys Pro Phe Ser Leu Thr Ile Ser
    290                 295                 300
Leu Asn Ser Ser Asn Tyr Ala Phe Met Gln Ala Leu Met His Ala Val
305                 310                 315                 320
Asp Pro Glu Ile Pro Gln Ala Val Cys Ile Pro Thr Lys Leu Ser Pro
                325                 330                 335
Ile Ser Met Leu Tyr Gln Asp Asn Asn Asp Asn Val Ile Leu Arg His
            340                 345                 350
Tyr Glu Asp Met Val Val Asp Glu Cys Gly Cys Gly
        355                 360
```

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1225 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:
(A) LIBRARY: HNT2NOT01
(B) CLONE: 266285

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

```
GTGGAGCTCT CCCCGGTCTG ACAGCCACTC CAGAGGCCAT GCTTCGTTTC TTGCCAGATT      60
TGGCTTTCAG CTTCCTGTTA ATTCTGGCTT TGGGCCAGGC AGTCCAATTT CAAGAATATG     120
TCTTTCTCCA ATTTCTGGGC TTAGATAAGG CGCCTTCACC CCAGAAGTTC CAACCTGTGC     180
CTTATATCTT GAAGAAAATT TTCCAGGATC GCGAGGCAGC AGCGACCACT GGGGTCTCCC     240
GAGACTTATG CTACGTAAAG GAGCTGGGCG TCCGCGGGAA TGTACTTCGC TTTCTCCCAG     300
ACCAAGGTTT CTTTCTTTAC CCAAAGAAAA TTTCCCAAGC TTCCTCCTGC CTGCAGAAGC     360
TCCTCTACTT TAACCTGTCT GCCATCAAAG AAAGGGAACA GTTGACATTG GCCCAGCTGG     420
GCCTGGACTT GGGGCCCAAT TCTTACTATA ACCTGGGACC AGAGCTGGAA CTGGCTCTGT     480
TCCTGGTTCA GGAGCCTCAT GTGTGGGGCC AGACCACCCC TAAGCCAGGT AAAATGTTTG     540
TGTTGCGGTC AGTCCCATGG CCACAAGGTG CTGTTCACTT CAACCTGCTG GATGTAGCTA     600
AGGATTGGAA TGACAACCCC CGGAAAAATT TCGGGTTATT CCTGGAGATA CTGGTCAAAG     660
AAGATAGAGA CTCAGGGGTG AATTTTCAGC CTGAAGACAA CTGTGCCAGA CTAAGATGCT     720
CCCTTCATGC TTCCCTGCTG GTGGTGACTC TCAACCCTGA TCAGTGCCAC CCTTCTCGGA     780
AAAGGAGAGC AGCCATCCCT GTCCCCAAGC TTTCTTGTAA GAACCTCTGC CACCGTCACC     840
AGCTATTCAT TAACTTCCGG GACCTGGGTT GGCACAAGTG GATCATTGCC CCCAAGGGGT     900
TCATGGCAAA TTACTGCCAT GGAGAGTGTC CCTTCTCACT GACCATCTCT CTCAACAGCT     960
CCAATTATGC TTTCATGCAA GCCCTGATGC ATGCCGTTGA CCCAGAGATC CCCCAGGCTG    1020
TGTGTATCCC CACCAAGCTG TCTCCCATTT CCATGCTCTA CCAGGACAAT AATGACAATG    1080
TCATTCTACG ACATTATGAA GACATGGTAG TCGATGAATG TGGGTGTGGG TAGGATGTCA    1140
GAAATGGGAA TAGAAGGAGT GTTCTTAGGG TAAATCTTTT AATAAAACTA CCTATCTGGT    1200
TTATGACCAC TTAGATCGAA ATGTC                                          1225
```

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 366 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:
(A) LIBRARY: HNT2NOT01
(B) CLONE: 15944

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

```
Met Gln Pro Tyr Gln Arg Leu Leu Ala Leu Gly Phe Leu Leu Leu Thr
 1               5                  10                  15
Leu Pro Trp Gly Gln Thr Ser Glu Phe Gln Asp Ser Asp Leu Leu Gln
            20                  25                  30
Phe Leu Gly Leu Glu Lys Ala Pro Ser Pro His Arg Phe Gln Pro Val
        35                  40                  45
Pro Arg Val Leu Arg Lys Ile Ile Arg Ala Arg Glu Ala Ala Ala Ala
    50                  55                  60
Ser Gly Ala Ser Gln Asp Leu Cys Tyr Val Lys Glu Leu Gly Val Arg
65                  70                  75                  80
Gly Asn Leu Leu Gln Leu Leu Pro Asp Gln Gly Phe Phe Leu Asn Thr
                85                  90                  95
Gln Lys Pro Phe Gln Asp Gly Ser Cys Leu Gln Lys Val Leu Tyr Phe
            100                 105                 110
Asn Leu Ser Ala Ile Lys Glu Lys Ala Lys Leu Thr Met Ala Gln Leu
        115                 120                 125
Thr Leu Asp Leu Gly Pro Arg Ser Tyr Tyr Asn Leu Arg Pro Glu Leu
    130                 135                 140
Val Val Ala Leu Ser Val Val Gln Asp Arg Gly Val Trp Gly Arg Ser
145                 150                 155                 160
His Pro Lys Val Gly Arg Leu Leu Phe Leu Arg Ser Val Pro Gly Pro
                165                 170                 175
```

```
Gln Gly Gln Leu Gln Phe Asn Leu Gln Gly Ala Leu Lys Asp Trp Ser
            180                 185                 190
Ser Asn Arg Leu Lys Asn Leu Asp Leu His Leu Glu Ile Leu Val Lys
        195                 200                 205
Glu Asp Arg Tyr Ser Arg Val Thr Val Gln Pro Glu Asn Pro Cys Asp
    210                 215                 220
Pro Leu Leu Arg Ser Leu His Ala Ser Leu Leu Val Val Thr Leu Asn
225                 230                 235                 240
Pro Lys His Cys His Pro Ser Ser Arg Lys Arg Arg Ala Ala Ile Ser
                245                 250                 255
Val Pro Lys Gly Phe Cys Arg Asn Phe Cys His Arg His Gln Leu Phe
            260                 265                 270
Ile Asn Phe Gln Asp Leu Gly Trp His Lys Trp Val Ile Ala Pro Lys
        275                 280                 285
Gly Phe Met Ala Asn Tyr Cys His Gly Glu Cys Pro Phe Ser Met Thr
    290                 295                 300
Thr Tyr Leu Asn Ser Ser Asn Tyr Ala Phe Met Gln Ala Leu Met His
305                 310                 315                 320
Met Ala Asp Pro Lys Val Pro Lys Ala Val Cys Val Pro Thr Lys Leu
                325                 330                 335
Ser Pro Ile Ser Met Leu Tyr Gln Asp Ser Asp Lys Asn Val Ile Leu
            340                 345                 350
Arg His Tyr Glu Asp Met Val Val Asp Glu Cys Gly Cys Gly
        355                 360                 365
```

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 366 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(vii) IMMEDIATE SOURCE:
(A) LIBRARY: GenBank
(B) CLONE: 567205

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

```
Met Gln Pro Tyr Gln Arg Leu Leu Ala Leu Gly Phe Leu Leu Leu Thr
 1               5                  10                  15
Leu Pro Trp Gly Gln Thr Ser Glu Phe Gln Asp Ser Asp Leu Leu Gln
            20                  25                  30
Phe Leu Gly Leu Glu Lys Ala Pro Ser Pro His Arg Phe Gln Pro Val
        35                  40                  45
Pro Arg Val Leu Arg Lys Ile Ile Arg Ala Arg Glu Ala Ala Ala Ala
    50                  55                  60
Ser Gly Ala Ser Gln Asp Leu Cys Tyr Val Lys Glu Leu Gly Val Arg
65                  70                  75                  80
Gly Asn Leu Leu Gln Leu Leu Pro Asp Gln Gly Phe Phe Leu Asn Thr
                85                  90                  95
Gln Lys Pro Phe Gln Asp Gly Ser Cys Leu Gln Lys Val Leu Tyr Phe
            100                 105                 110
Asn Leu Ser Ala Ile Lys Glu Lys Ala Lys Leu Thr Met Ala Gln Leu
        115                 120                 125
Thr Leu Asp Leu Gly Pro Arg Ser Tyr Tyr Asn Leu Arg Pro Glu Leu
    130                 135                 140
Val Val Ala Leu Ser Val Val Gln Asp Arg Gly Val Trp Gly Arg Ser
145                 150                 155                 160
```

```
His Pro Lys Val Gly Arg Leu Leu Phe Leu Arg Ser Val Pro Gly Pro
                165                 170                 175
Gln Gly Gln Leu Gln Phe Asn Leu Gln Gly Ala Leu Lys Asp Trp Ser
            180                 185                 190
Ser Asn Arg Leu Lys Asn Leu Asp Leu His Leu Glu Ile Leu Val Lys
        195                 200                 205
Glu Asp Arg Tyr Ser Arg Val Thr Val Gln Pro Glu Asn Pro Cys Asp
    210                 215                 220
Pro Leu Leu Arg Ser Leu His Ala Ser Leu Leu Val Val Thr Leu Asn
225                 230                 235                 240
Pro Lys His Cys His Pro Ser Ser Arg Lys Arg Arg Ala Ala Ile Ser
                245                 250                 255
Val Pro Lys Gly Phe Cys Arg Asn Phe Cys His Arg His Gln Leu Phe
            260                 265                 270
Ile Asn Phe Gln Asp Leu Gly Trp His Lys Trp Val Ile Ala Pro Lys
        275                 280                 285
Gly Phe Met Ala Asn Tyr Cys His Gly Glu Cys Pro Phe Ser Met Thr
    290                 295                 300
Thr Tyr Leu Asn Ser Ser Asn Tyr Ala Phe Met Gln Ala Leu Met His
305                 310                 315                 320
Met Ala Asp Pro Lys Val Pro Lys Ala Val Cys Val Pro Thr Lys Leu
                325                 330                 335
Ser Pro Ile Ser Met Leu Tyr Gln Asp Ser Asp Lys Asn Val Ile Leu
            340                 345                 350
Arg His Tyr Glu Asp Met Val Val Asp Glu Cys Gly Cys Gly
        355                 360                 365
```

What is claimed is:

1. A substantially purified human growth and differentiation protein comprising the amino acid sequence of SEQ ID NO:1 or fragments thereof.

2. An isolated and purified polynucleotide sequence encoding the human growth and differentiation protein of claim 1.

3. A polynucleotide sequence which hybridizes under stringent conditions to the polynucleotide sequence of claim 2.

4. A hybridization probe comprising the polynucleotide sequence of claim 2.

5. An isolated and purified polynucleotide sequence comprising SEQ ID NO:2 or variants thereof.

6. A polynucleotide sequence which is complementary to the polynucleotide sequence of claim 2 or variants thereof.

7. A hybridization probe comprising the polynucleotide sequence of claim 6.

8. An expression vector containing the polynucleotide sequence of claim 2.

9. A host cell containing the vector of claim 8.

10. A method for producing a polypeptide comprising the amino acid sequence of SEQ ID NO:1 the method comprising the steps of:

a) culturing the host cell of claim 9 under conditions suitable for the expression of the polypeptide; and

b) recovering the polypeptide from the host cell culture.

11. A pharmaceutical composition comprising a substantially purified human growth and differentiation protein having an amino acid sequence of SEQ ID NO:1 in conjunction with a suitable pharmaceutical carrier.

12. A purified antibody which binds specifically to the polypeptide of claim 1.

13. A purified agonist which modulates the activity of the polypeptide of claim 1.

14. A purified antagonist which modulates the activity of the polypeptide of claim 1.

15. A pharmaceutical composition comprising an antagonist of claim 14 in conjunction with a suitable pharmaceutical carrier.

16. A method for stimulating cell proliferation comprising adding to a cell an effective amount of the protein of claim 1.

17. A method for treating or preventing a disorder associated with cell proliferation comprising administering to a subject in need of such treatment the pharmaceutical composition of claim 15.

18. A method for detection of polynucleotides encoding human growth and differentiation protein in a biological sample comprising the steps of:

a) hybridizing the polynucleotide of claim 6 to nucleic acid material of a biological sample, thereby forming a hybridization complex; and

b) detecting said hybridization complex, wherein the presence of said complex correlates with the presence of a polynucleotide encoding human growth and differentiation protein in said biological sample.

```
          9          18          27          36          45          54
 GT GGA GCT CTC CCC GGT CTG ACA GCC ACT CCA GAG GCC ATG CTT CGT TTC TTG
                                                   M   L   R   F   L

         63          72          81          90          99         108
CCA GAT TTG GCT TTC AGC TTC CTG TTA ATT CTG GCT TTG GGC CAG GCA GTC CAA
P   D   L   A   F   S   F   L   L   I   L   A   L   G   Q   A   V   Q

        117         126         135         144         153         162
TTT CAA GAA TAT GTC TTT CTC CAA TTT CTG GGC TTA GAT AAG GCG CCT TCA CCC
F   Q   E   Y   V   F   L   Q   F   L   G   L   D   K   A   P   S   P

        171         180         189         198         207         216
CAG AAG TTC CAA CCT GTG CCT TAT ATC TTG AAG AAA ATT TTC CAG GAT CGC GAG
Q   K   F   Q   P   V   P   Y   I   L   K   K   I   F   Q   D   R   E

        225         234         243         252         261         270
GCA GCA GCG ACC ACT GGG GTC TCC CGA GAC TTA TGC TAC GTA AAG GAG CTG GGC
A   A   A   T   T   G   V   S   R   D   L   C   Y   V   K   E   L   G

        279         288         297         306         315         324
GTC CGC GGG AAT GTA CTT CGC TTT CTC CCA GAC CAA GGT TTC TTT CTT TAC CCA
V   R   G   N   V   L   R   F   L   P   D   Q   G   F   F   L   Y   P

        333         342         351         360         369         378
AAG AAA ATT TCC CAA GCT TCC TCC TGC CTG CAG AAG CTC CTC TAC TTT AAC CTG
K   K   I   S   Q   A   S   S   C   L   Q   K   L   L   Y   F   N   L
```

FIGURE 1A

```
          387        396        405        414        423        432
TCT GCC ATC AAA GAA AGG GAA CAG TTG ACA TTG GCC CAG CTG GGC CTG GAC TTG
S   A   I   K   E   R   E   Q   L   T   L   A   Q   L   G   L   D   L

          441        450        459        468        477        486
GGG CCC AAT TCT TAC TAT AAC CTG GGA CCA GAG CTG GAA CTG GCT CTG TTC CTG
G   P   N   S   Y   Y   N   L   G   P   E   L   E   L   A   L   F   L

          495        504        513        522        531        540
GTT CAG GAG CCT CAT GTG TGG GGC CAG ACC ACC CCT AAG CCA GGT AAA ATG TTT
V   Q   E   P   H   V   W   G   Q   T   T   P   K   P   G   K   M   F

          549        558        567        576        585        594
GTG TTG CGG TCA GTC CCA TGG CCA CAA GGT GCT GTT CAC TTC AAC CTG CTG GAT
V   L   R   S   V   P   W   P   Q   G   A   V   H   F   N   L   L   D

          603        612        621        630        639        648
GTA GCT AAG GAT TGG AAT GAC AAC CCC CGG AAA AAT TTC GGG TTA TTC CTG GAG
V   A   K   D   W   N   D   N   P   R   K   N   F   G   L   F   L   E

          657        666        675        684        693        702
ATA CTG GTC AAA GAA GAT AGA GAC TCA GGG GTG AAT TTT CAG CCT GAA GAC AAC
I   L   V   K   E   D   R   D   S   G   V   N   F   Q   P   E   D   N

          711        720        729        738        747        756
TGT GCC AGA CTA AGA TGC TCC CTT CAT GCT TCC CTG CTG GTG GTG ACT CTC AAC
C   A   R   L   R   C   S   L   H   A   S   L   L   V   V   T   L   N
```

FIGURE 1B

```
          765         774         783         792         801         810
CCT GAT CAG TGC CAC CCT TCT CGG AAA AGG AGA GCA GCC ATC CCT GTC CCC AAG
P   D   Q   C   H   P   S   R   K   R   R   A   A   I   P   V   P   K

          819         828         837         846         855         864
CTT TCT TGT AAG AAC CTC TGC CAC CGT CAC CAG CTA TTC ATT AAC TTC CGG GAC
L   S   C   K   N   L   C   H   R   H   Q   L   F   I   N   F   R   D

          873         882         891         900         909         918
CTG GGT TGG CAC AAG TGG ATC ATT GCC CCC AAG GGG TTC ATG GCA AAT TAC TGC
L   G   W   H   K   W   I   I   A   P   K   G   F   M   A   N   Y   C

          927         936         945         954         963         972
CAT GGA GAG TGT CCC TTC TCA CTG ACC ATC TCT CTC AAC AGC TCC AAT TAT GCT
H   G   E   C   P   F   S   L   T   I   S   L   N   S   S   N   Y   A

          981         990         999        1008        1017        1026
TTC ATG CAA GCC CTG ATG CAT GCC GTT GAC CCA GAG ATC CCC CAG GCT GTG TGT
F   M   Q   A   L   M   H   A   V   D   P   E   I   P   Q   A   V   C

         1035        1044        1053        1062        1071        1080
ATC CCC ACC AAG CTG TCT CCC ATT TCC ATG CTC TAC CAG GAC AAT AAT GAC AAT
I   P   T   K   L   S   P   I   S   M   L   Y   Q   D   N   N   D   N

         1089        1098        1107        1116        1125        1134
GTC ATT CTA CGA CAT TAT GAA GAC ATG GTA GTC GAT GAA TGT GGG TGT GGG TAG
V   I   L   R   H   Y   E   D   M   V   V   D   E   C   G   C   G
```

FIGURE 1C

```
        1143        1152        1161        1170        1179        1188
GAT GTC AGA AAT GGG AAT AGA AGG AGT GTT CTT AGG GTA AAT CTT TTA ATA AAA

        1197        1206        1215        1224
CTA CCT ATC TGG TTT ATG ACC ACT TAG ATC GAA ATG TC
```

FIGURE 1D

```
1    MLRFLPDLAFSFLLILALGQAVQFQEYVFL 266285
1    MQPYQRLLALGFLLLTLPWGQTSEFQDSDL WO94/15944
1    MQPYQRLLALGFLLLTLPWGQTSEFQDSDL GI 567205

31   QFLGLDKAPSPQKFQPVPYILKKIFQDREA 266285
31   LQFLGLEKAPSPHRFQPVPRVLRKIIRARE WO94/15944
31   LQFLGLEKAPSPHRFQPVPRVLRKIIRARE GI 567205

61   AATTGVSRDLCYVKELGVRGNVLRFLPDQG 266285
61   AAAASGASQDLCYVKELGVRGNLLQLLPDQ WO94/15944
61   AAAASGASQDLCYVKELGVRGNLLQLLPDQ GI 567205

91   FFLYPKKISQASSCLQKLLYFNLSAIKERE 266285
91   GFFLNTQKPFQDGSCLQKVLYFNLSAIKEK WO94/15944
91   GFFLNTQKPFQDGSCLQKVLYFNLSAIKEK GI 567205

121  QLTLAQLGLDLGPNSYYNLGPELELALFLV 266285
121  AKLTMAQLTLDLGPRSYYNLRPELVVALSV WO94/15944
121  AKLTMAQLTLDLGPRSYYNLRPELVVALSV GI 567205

151  QEPHVWGQTTPKPGKMFVLRSVPWPQGAVH 266285
151  VQDRGVWGRSHPKVGRLLFLRSVPGPQGQL WO94/15944
151  VQDRGVWGRSHPKVGRLLFLRSVPGPQGQL GI 567205
```

FIGURE 2A

```
181 FNLLDVAKDWNDNPRKNFGLFLEILVKEDR 266285
181 QFNLQGALKDWSSNRLKNLDLHLEILVKED WO94/15944
181 QFNLQGALKDWSSNRLKNLDLHLEILVKED GI 567205

211 DSGVNFQPEDNCARLRCSLHASLLVVTLNP 266285
211 RYSRVTVQPENPCDPLLRSLHASLLVVTLN WO94/15944
211 RYSRVTVQPENPCDPLLRSLHASLLVVTLN GI 567205

241 DQCHPSRKRRAAIPVPKLSCKNLCHRHQLF 266285
241 PKHCHPSSRKRRAAISVPKGFCRNFCHRHQ WO94/15944
241 PKHCHPSSRKRRAAISVPKGFCRNFCHRHQ GI 567205

271 INFRDLGWHKWIIAPKGFMANYCHGECPFS 266285
271 LFINFQDLGWHKWVIAPKGFMANYCHGECP WO94/15944
271 LFINFQDLGWHKWVIAPKGFMANYCHGECP GI 567205

301 LTISLNSSNYAFMQALMHAVDPEIPQAVCI 266285
301 FSMTTYLNSSNYAFMQALMHMADPKVPKAV WO94/15944
301 FSMTTYLNSSNYAFMQALMHMADPKVPKAV GI 567205

331 PTKLSPISMLYQDNNDNVILRHYEDMVVDE 266285
331 CVPTKLSPISMLYQDSDKNVILRHYEDMVV WO94/15944
331 CVPTKLSPISMLYQDSDKNVILRHYEDMVV GI 567205
```

FIGURE 2B

266285
WO94/15944
GI 567205

361 C G C G C G C G
361 D E C G C G C G
361 D E C G C G C G

FIGURE 2C

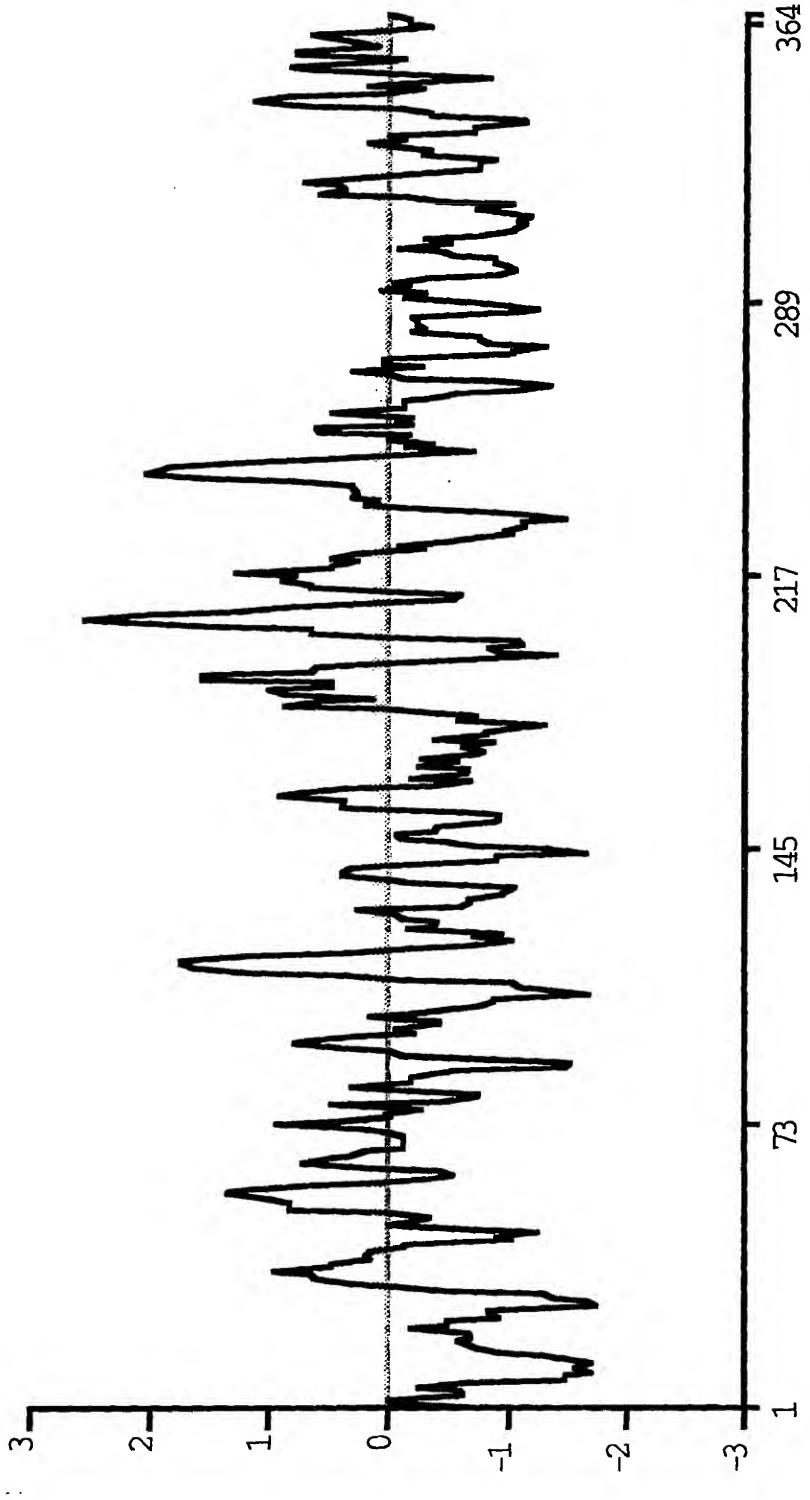

FIGURE 3A

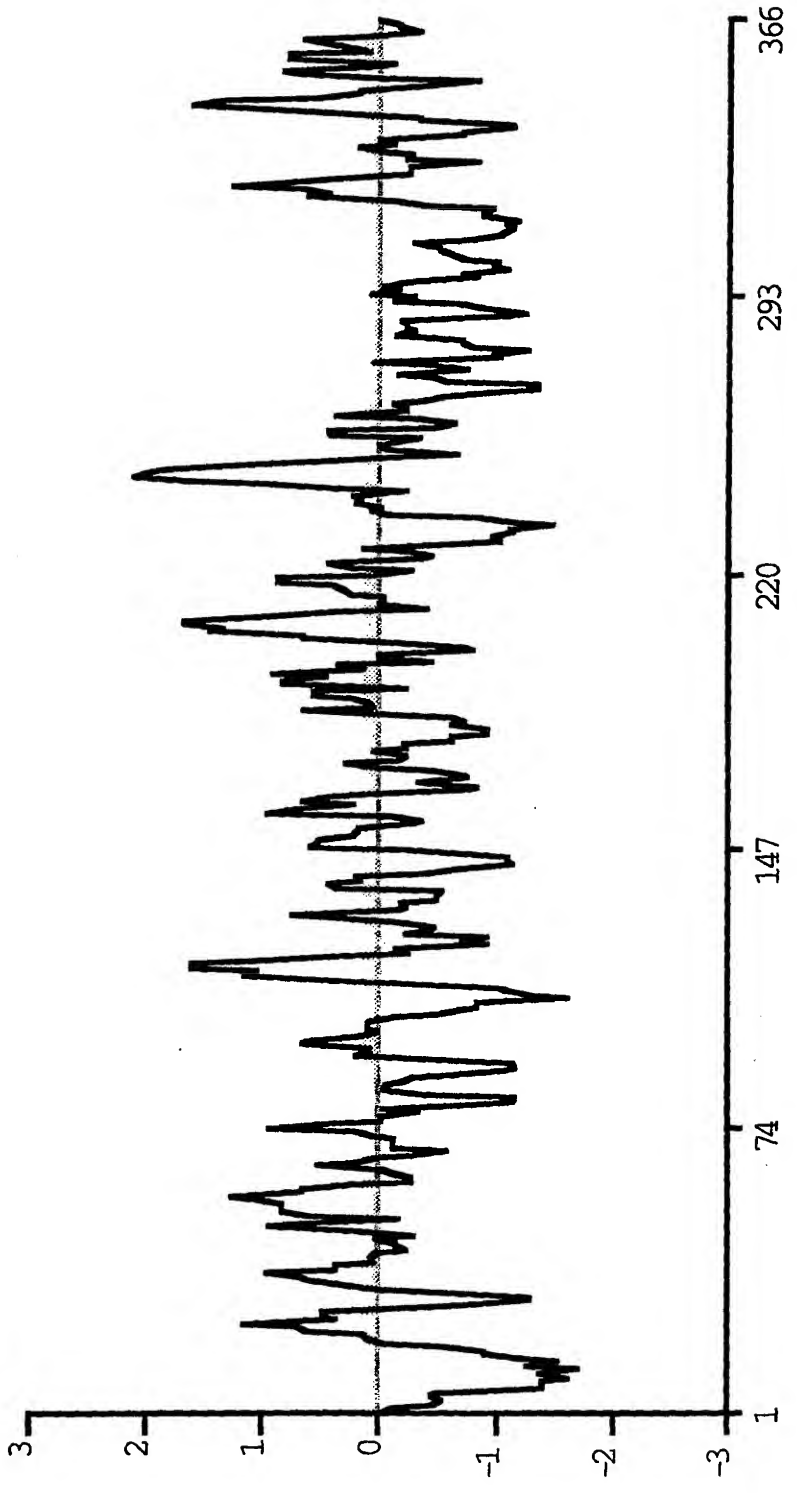

FIGURE 3B

INTERNATIONAL SEARCH REPORT

International Application No
PCT/US 98/05700

A. CLASSIFICATION OF SUBJECT MATTER
IPC 6 C12N15/12 C07K14/475 A61K38/18 C12Q1/68 C07K16/22
//C12N15/62,C12N9/00,C12N15/11

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 6 C12N C07K A61K C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	WO 94 15965 A (UNIV JOHNS HOPKINS MED ;LEE SE JIN (US); MCPHERRON ALEXANDRA C (US) 21 July 1994 see the whole document, especially seq. 28, examples 1-3 and claims 1,2,5,8,11,23 ---	1-4, 6-12, 14-18
X	DATABASE EMBL - EMEST4 Entry HS1136360, Acc.No. AA218985, 12 February 1997 HILLIER, L. ET AL.: "zr01f06.s1 Stratagene NT2 neuronal precursor 937230 Homo sapiens cDNA clone 650243 3' similar to SW:GDF3_MOUSE Q07104 GROWTH/DIFFERENTIATION FACTOR 3 PRECURSOR." XP002068140 see the whole document --- -/--	1-10,18

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
19 June 1998	09.07.98

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL - 2280 HV Rijswijk Tel. (+31-70) 340-2040, Tx. 31 651 epo nl, Fax: (+31-70) 340-3016	Smalt, R

Form PCT/ISA/210 (second sheet) (July 1992)

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	DATABASE EMBL - R54U006 Entry HSAA6111, Acc.No. AA206111, 31 January 1997 HILLIER, L. ET AL.: "zq52g05.r1 Stratagene neuroepithelium (#937231) Homo sapiens cDNA clone 645272 5' similar to SW:GDF3_MOUSE Q07104 GROWTH/DIFFERENTIATION FACTOR 3 PRECURSOR." XP002068141 see the whole document ---	1-10,18
P,X	WO 97 35870 A (HUMAN GENOME SCIENCES INC ;SOPPET DANIEL R (US); LI HAODONG (US)) 2 October 1997 see page 18, paragraph 3; claims 16-18,20,22 -----	1-18

1

Form PCT/ISA/210 (continuation of second sheet) (July 1992)

INTERNATIONAL SEARCH REPORT

tional application No.
PCT/US 98/05700

Box I Observations where certain claims wer found unsearchable (Continuati n of item 1 f first sheet)

This International Search Report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. [X] Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

Although claims 17, and 16 in as far as it relates to in vivo use, are directed to a method of treatment of the human/animal body, the search has been carried out and based on the alleged effects of the compound/composition.

2. [ ] Claims Nos.:
because they relate to parts of the International Application that do not comply with the prescribed requirements to such an extent that no meaningful International Search can be carried out, specifically:

3. [ ] Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. [ ] As all required additional search fees were timely paid by the applicant, this International Search Report covers all searchable claims.

2. [ ] As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.

3. [ ] As only some of the required additional search fees were timely paid by the applicant, this International Search Report covers only those claims for which fees were paid, specifically claims Nos.:

4. [ ] No required additional search fees were timely paid by the applicant. Consequently, this International Search Report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

[ ] The additional search fees were accompanied by the applicant's protest.

[ ] No protest accompanied the payment of additional search fees.

INTERNATIONAL SEARCH REPORT

Information on patent family members

Inte[redacted] Application No
PC[redacted]S 98/05700

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 9415965 A	21-07-1994	CA 2153652 A	21-07-1994
		EP 0679163 A	02-11-1995
		JP 8505771 T	25-06-1996
WO 9735870 A	02-10-1997	AU 5527296 A	17-10-1997

Form PCT/ISA/210 (patent family annex) (July 1992)